



NATO
|
OTAN

Headquarters Supreme Allied Commander

Transformation

Autonomous Systems

Issues for Defence
Policymakers



Edited by:

Andrew P. Williams

Paul D. Scharre

Autonomous Systems

Issues for Defence Policymakers

Autonomous Systems

Issues for Defence Policymakers

Edited by:

Andrew P. Williams

Paul D. Scharre



Capability Engineering and Innovation Division
Headquarters Supreme Allied Commander Transformation
Norfolk, Virginia
United States



Capability Engineering and Innovation
HQ SACT, Suite 100
7857 Blandy Road
Norfolk, VA 23551
United States
<http://www.act.nato.int>

Portions of this work may be quoted without permission, provided that standard citations practices are used. Copyright is retained by the authors, or their respective parent organizations where indicated in each chapter.

This book is also available on line at <http://www.act.nato.int>

ISBN 9789284501939

Opinions, conclusions, and recommendations expressed or implied within are solely those of the contributors and do not necessarily represent the views of NATO, Headquarters Supreme Allied Commander Transformation, or any of the author's respective parent organisations. All authors are responsible for their own content. NATO, Headquarters Supreme Allied Commander Transformation, or the Editors, are not responsible for any omissions, errors, or false statements.

Printed by:
NATO Communications and Information Agency,
Oude Waalsdorperweg 61,
2597 AK The Hague, Netherlands



Contents

Foreword - <i>General Jean-Paul Paloméros, Supreme Allied Commander Transformation</i>	iii
Preface - <i>Brigadier General Henrik Sommer, Assistant Chief of Staff, Capability Engineering and Innovation, HQ SACT</i>	iv
Acknowledgements - <i>The Editors</i>	vii
Contributors	viii
PART 1: INTRODUCTION	
Chapter 1	3
The Opportunity and Challenge of Autonomous Systems	
<i>Paul Scharre</i>	
Chapter 2	27
Defining Autonomy in Systems: Challenges and Solutions	
<i>Andrew Williams</i>	
PART 2: ETHICAL, LEGAL, AND POLICY PERSPECTIVES	
Chapter 3	65
Developing Autonomous Systems in an Ethical Manner	
<i>Chris Mayer</i>	
Chapter 4	83
The Legal Implications of the Use of Systems with Autonomous Capabilities in Military Operations	
<i>Roberta Arnold</i>	
Chapter 5	98
The Varied Law of Autonomous Weapon Systems	
<i>Rebecca Crootoof</i>	
Chapter 6	127
Civilian Developments in Autonomous Vehicle Technology and their Civilian and Military Policy Implications	
<i>James M. Anderson & John Matsumura</i>	

PART 3: AUTONOMOUS SYSTEMS AND OPERATIONAL RISK

Chapter 7	152
NATO's Targeting Process: Ensuring Human Control Over (and Lawful Use of) 'Autonomous' Weapons	
<i>Mark Roorda</i>	
Chapter 8	169
Choosing the Level of Autonomy: Options and Constraints	
<i>Erik Theunissen & Brandon Suarez</i>	
Chapter 9	196
Auditable Policies for Autonomous Systems (Decisional Forensics)	
<i>Tom Keeley</i>	

PART 4: PERSPECTIVES ON IMPLEMENTING AUTONOMY IN SYSTEMS

Chapter 10	226
Four Challenges in Structuring Human-Autonomous Systems Interaction Design Processes	
<i>Pertti Saariluoma</i>	
Chapter 11	249
Passive Optical Sensor Systems for Navigation in Unstructured, Non-cooperating Environments	
<i>Christoph Sulzbachner, Christian Zinner, & Thomas Kadiofsky</i>	
Chapter 12	263
Toward Defining Canadian Manned-Unmanned Teaming (MUM-T) Concepts	
<i>Benoit Arbour, Matthew R. MacLeod & Sean Bourdon</i>	
Chapter 13	291
Merging Two Worlds: Agent-based Simulation Methods for Autonomous Systems	
<i>Andreas Tolk</i>	
Afterword	318
Looking Forward – A Research Agenda for NATO	
<i>Major-General Husniaux, NATO Chief Scientist</i>	

Foreword

I am pleased to introduce the second volume of Allied Command Transformation's *Innovation in Capability Development* series. I launched this initiative to provide a forum for the study and exploration of innovative ideas and concepts, in order to contribute to NATO capability development.

This volume covers the subject of autonomous systems, which stand potentially to transform the way in which warfare is conducted. Advances in sensors, robotics and computing are permitting the development of a whole new class of systems, which offer a wide range of military benefits including the ability to operate without personnel on board the platform, novel human-machine teaming concepts, and “swarm” operating methods. With any such transformational advances, there are unique operational, legal, ethical and design issues. It is critical that we engage in open and transparent debate about the wide ranging opportunities, implications and challenges that these advances will bring, and ensure that the Alliance is ready to address the forthcoming issues of terminology, standardisation, interoperability and military operational conduct.

This work had its genesis in the 2013-14 Multinational Capability Development Campaign—a collaborative programme of work between 19 nations, NATO and the European Union. Headquarters Supreme Allied Commander Transformation led the autonomous systems focus area, which brought together a group of scientific, policy, legal and military experts to study this critical defence policy issue. This volume elaborates more on their contributions, in addition to welcoming several new authors. Their studies are presented in a high quality, peer-reviewed and edited volume, which is released publicly to encourage wide readership and debate.

I believe this volume is a noteworthy contribution to the current debate on autonomous systems in our Alliance and partner nations, and is unique in presenting a multidisciplinary view of the topic. There are few issues which are as significant to capability development as the rise of unmanned and autonomous systems and their effect on the principles of warfare; this volume takes an important step in clarifying our understanding of the implications.

Jean-Paul Paloméros,
General, French Air Force
Supreme Allied Commander Transformation



Preface

It is an enormous honour to open the second volume in the *Innovation in Capability Development* publication series—*Autonomous Systems: Issues for Defence Policymakers*. A core task of my Capability Engineering and Innovation Division at Headquarters Supreme Allied Commander Transformation (HQ SACT) is to drive innovation in NATO by overcoming organizational constraints, challenging established ways of working and thinking, and taking a long term perspective on what change is possible. There are many tools for accomplishing innovation, including studies, concept development and experimentation, networking with experts from nations, academia, and industry, and collaborative online venues such as the Innovation Hub.

This series of edited volumes is another vector for innovation in NATO. The first volume in the series—*Innovation in Operations Assessment*—was highly successful. The book has been downloaded almost 4000 times from the HQ SACT website, is required reading for several NATO and national training courses, and recently was influential in the development of US doctrine on operations assessment. I have no doubt that this second volume will be as equally successful, especially given the subject.

The increasingly autonomous capabilities of military systems raise a variety of challenging legal, ethical, policy, operational, and technical issues. The major contribution of this volume lies in its multidisciplinary presentation of these challenges. In the introduction section, Scharre outlines the basic concept of autonomy and reflects on its impact on the conduct of military operations such as the opportunity for human-machine teaming. Williams then presents a detailed analysis on how “autonomy” is defined. The chapters by Scharre and Williams remind us that truly autonomous systems do not exist—except perhaps in the distant futures of science fiction stories. Instead, they stress that autonomy is a capability of a system, which exists in a complex arrangement of organizations, policies, guidance and human control.

The chapter by Roorda gives a concrete example of how the use of autonomous systems is governed by specific policy processes, and how each stage of the process incorporates the necessary human decision-making over relevant legal, ethical and operational questions, before any system is deployed.

Yet contemporary policy debate—most recently at the United Nations Convention on Certain Conventional Weapons—shows that autonomous systems remain a contentious subject. While the Law Of Armed Conflict and International Human Rights Law do not prohibit delegation of military functions to autonomous systems in general, the chapters by Arnold and Crooto broaden the picture and discuss, in particular, the roles and responsibilities of states, and the wide variety of international law that may be affected by the development of autonomous systems. Such policy issues are not unique to the military domain, however. The chapter by Anderson and Matsumura highlights legal and policy challenges facing the introduction of “driverless” cars in the civilian world, of which military capability should take note.

Even in the light of these legal and policy conclusions, serious ethical and practical reservations about autonomous systems remain. The chapter by Mayer lays out an ethical framework for analysing the implications of autonomous systems in military operations and presents a series of policy-maker recommendations. Keeley offers another solution, suggesting a process and capability to audit and trace the internal algorithms of systems for the purposes of verification and validation, testing and accountability. Theunissen and Suarez show, however, that for many tasks high levels of autonomy are not necessarily desirable and that optimal solutions involve combining autonomous and human control to leverage the benefits of each.

With this legal, definitional and ethical framework for policy considerations of autonomous systems, the last section of the book presents four, detailed chapters on various aspects of capability development. First, Saariluoma presents a framework for introducing human-technology interaction design-based thinking for autonomous systems, reminding us that ultimately, such systems are human tools and therefore the way in which humans interact and control the system should be intuitive.

The chapter by Sulzbachner, Zinner and Kadiofsky emphasizes that a requirement for many autonomous systems is passive optical sensor technology for motion in unstructured, non-cooperating environments. They remind us that the various regulatory initiatives ongoing to incorporate unmanned and autonomous systems into airspace, for example, rely inherently on establishing performance standards, yet standardised approaches for verifying optical and other sensor systems are still not fully established.

Arbour, MacLeod and Bourdon offer an analysis of manned-unmanned teaming concepts and present a range of different options of how unmanned aircraft can be used in concert with manned aircraft for a variety of novel roles. Their advice is that, given the increasing reliability and autonomy of unmanned systems, we need to define a greater number of potential configurations and combinations in the early stages of capability development, rather than be limited to one-to-one platform replacements.

The final chapter of the volume by Tolk recommends the increased use of agent-based simulation methods to support the design, development, testing, and operational use of autonomous systems. The reason is that software agents in a simulation are the virtual counterparts of autonomous robotic systems, and in many cases, the underlying algorithms governing physical system and software agent may be the same. This leads to many opportunities to improve system test and evaluation, but also operational support by running a system through simulations for its mission tasks before deployment.

The volume closes with a forward thinking view from the NATO Chief Scientist, Major-General Husniaux, who recommends a series of key research areas for future capability development. He reminds us that science and technology development is essential, but must co-evolve in collaboration with operator, policy, legal and doctrinal views.

We are expanding our horizons in this critical area defence capability, yet innovation requires outside input and fresh perspectives. The best way to achieve this is by sharing ideas through a diverse network of experts. I know that the work led by HQ SACT has benefitted tremendously from this exposure, and I hope that the Alliance network of scientists, engineers, legal and policy staffs, and military operators will benefit likewise.

A handwritten signature in black ink, appearing to read 'H. Sommer', with a long horizontal flourish extending to the right.

*Brigadier General Henrik Sommer,
Danish Army*

*Assistant Chief of Staff, Capability Engineering and Innovation,
Headquarters Supreme Allied Commander Transformation*

Acknowledgements

There are many people who worked hard to make this volume happen. First, we are indebted to the authors who provided their contributions at no cost to NATO and were patient enough to deal with changing timelines, multiple revisions and reviews. It has been an enlightening experience to work with such a diverse group of experts and the results of their efforts is a work that is at the forefront of current analysis in this field. Second, we are thankful to the external reviewers who spent much time carefully scrutinizing the draft articles and providing many helpful suggestions, and to the original members of the Multinational Capability Development Campaign autonomous systems focus area, who supported the idea of this volume from the outset. Many thanks to Dr Alex Bourque, Mr John Canning, Mr Harry Drottz, Dr Candace Eschelman-Haynes, Lt Col Sylwester Filipczak, Ms. Candace Hetzer, Professor Patrick Lin, Mr Simon Purton, Lt Col Jeffrey Thurner, Capt Stefan Trnka, and Mr Larrel Walters. Third, we want to thank the NATO Communications and Information Agency for preparing the manuscript and printing the volume.

The Editors

Contributors

James M. Anderson is a senior behavioural scientist at the RAND Corporation and a professor at the Pardee RAND Graduate School. His current research focuses primarily on civil and criminal law. His projects include improving the use of forensic evidence, the effect of regulating medical marijuana, and the policy implications of autonomous vehicle technology. Among other topics, Anderson has published on the effect of zoning on crime, the indeterminacy of the economic analysis of tort law, no-fault automobile insurance, and the liability implications of autonomous vehicles. Before joining RAND, he practiced law for ten years. Anderson received a J.D. from Yale Law School and a B.A. in Ethics, Politics, and Economics from Yale University.

Benoit Arbour is a Defence Scientist with Defence Research and Development Canada, Centre for Operational Research and Analysis, currently assigned to the Maritime Operational Research Team in Ottawa, Ontario. His involvement in unmanned systems includes working on the introduction of the Scan Eagle unmanned aerial vehicle onto Royal Canadian Navy warships, studying unmanned aircraft usage in a maritime force protection role, assessing speed and endurance trade-offs in domestic operations for Canada's Joint Unmanned Surveillance Target Acquisition System, and being co-leader of the multi-research centre Manned-Unmanned Aerial Vehicle Interactions study. He holds a BMath in Pure and Applied Mathematics from the University of Waterloo and an M.Sc. in Number Theory from the University of McGill.

Roberta Arnold is a military investigating magistrate (specialist officer/captain) within the Military Justice of the Swiss Armed Forces. She has previously held different functions within the Swiss Federal Department of Defence (DoD). Between 2012 and 2015 she worked as researcher for the Chair of Strategic Studies of the Military Academy at the Swiss Federal Institute of Technology (ETH Zurich); at the same time she

held a part-time position as lawyer within the Center of Competence for International Criminal Law of the Federal Attorney General's Office. Previously, she was a legal adviser on the Laws of Armed Conflict and a policy advisor on arms control and disarmament policy within the Division of International Relations of the Swiss DoD. She has been invited to give presentations and lectures on the Laws of Armed Conflict and International Criminal Law at different academic institutions, including the Universities of Salzburg, Trento, Bern, Zurich, Liverpool, Bristol, Stanford, Pittsburgh and the G.C. Marshall Centre in Garmisch-Partenkirchen, amongst others. She is a member of the International Society for Military Law and the Laws of War, where she currently acts as Editor-in-chief of the Military Law and the Laws of War Review (www.mllwr.org). She holds a Ph.D. from the University of Bern and an LL.M from the University of Nottingham.

Sean Bourdon has been a Defence Scientist with Defence Research and Development Canada, Centre for Operational Research and Analysis since 2002. His current assignment is as Head of the Maritime and Air Systems Operational Research Section, whose mandate is to support both the Royal Canadian Navy and the Royal Canadian Air Force. His involvement with unmanned aerial vehicles came as a result of two previous assignments with the Royal Canadian Air Force's headquarters staff, and included an evaluation of the potential utility of unmanned aircraft for domestic maritime roles in addition to the present work. He holds BMath and MMath degrees in Applied Math from the University of Waterloo, Canada.

Rebecca Crotoft is pursuing a Ph.D. in Law at the Yale Graduate School of Arts and Sciences and is a resident fellow with the Yale Law School Information Society Project. Her research focuses on how international law evolves and its role in American law and policy. Her primary areas of interest include the law of armed conflict, human rights law, and the interplay between law and disruptive technologies. Crotoft earned her B.A. cum laude at Pomona College and her J.D. at Yale Law School, where she was an active member of the Center for Global Legal

Challenges. Prior to returning to Yale, Crotoft served as a law clerk for Judge Mark R. Kravitz of the US District Court for the District of Connecticut and for Judge John M. Walker, Jr. of the US Court of Appeals for the Second Circuit.

Thomas Kadiofsky graduated in Mathematical Computer Science at Vienna University of Technology in 2011. He joined the AIT Austrian Institute of Technology in 2009. Since then he has been working in the area of computer vision with a special focus on real-time 3D modelling.

Tom Keeley is the president of Compsim LLC and the inventor of KEEL® Technology. He received his B.S. degree from the College of Engineering and Technology of Bradley University. He had experience in software development, technical management, and business and technical strategy development at General Electric and Rockwell International before starting Compsim in 1999. His fields of endeavour have included inertial guidance and fire control systems, data communication systems (fixed and mobile), industrial robotics, intelligent sensor systems, and distributed system architectures. Recently he has focused on knowledge capture and decision-making techniques that allow devices to exercise judgement and reason on their own.

Matthew R. MacLeod is a Defence Scientist with Defence Research and Development Canada, Centre for Operational Research and Analysis, currently embedded in a small team within the Canadian Forces Maritime Warfare Centre in Halifax, Nova Scotia. In addition to the work described herein, his experience with unmanned systems includes analysis support to a workshop on potential adversary use of maritime remote autonomous systems, and joining the autonomous systems activity within The Technical Cooperation Panel's Maritime Technical Panel 3 in spring 2014. His Master's work (MASc, Carleton University) focused on heuristics for non-linear optimisation problems, and his work experience has primarily been in the areas of decision support and operational analysis of human-in-the-loop simulations. He recently completed a four month operational deployment based

in Bahrain with the Combined Maritime Forces.

John Matsumura is a senior engineer at the RAND Corporation with experience conducting research on a wide range of advanced technologies. His recent research includes autonomous robotics, force protection, operational energy, and systems-level analyses. In addition to serving in a number of research management roles at RAND, Matsumura served on government advisory committees, including: member of the Defense Science Board task force on power projection, an Office of the Secretary of Defense task force on defence architecture, and the Army Science Board, where he chaired a variety of panels and studies. He is a recipient of the RAND President's Award, and he recently received the Commander's Award for outstanding civilian service from the Office of the Assistant Secretary of the Army, Acquisition, Logistics, and Technology. In addition to his work at RAND, Matsumura is currently an adjunct faculty member at the Carnegie Institute of Technology, within Carnegie Mellon University's school of engineering, where he teaches courses at the graduate level. Matsumura received B.S. and M.S. degrees in Aerospace Engineering and Engineering Mechanics from Pennsylvania State University, and a Ph.D. in Engineering and Public Policy from Carnegie Mellon University.

Chris Mayer is an assistant professor of philosophy in the Department of English and Philosophy at the United States Military Academy (West Point) and a lieutenant colonel in the US Army. During his time at West Point, he has served as director of the academy's introduction to philosophy course, which is a graduation requirement for all cadets. He has also served in a variety of command and staff positions during his 21 years in the Army. His research focuses on moral philosophy, just war theory, and political philosophy.

Mark Roorda is pursuing a doctoral degree in law at the University of Amsterdam and the Netherlands Defence Academy. His research focuses on the legitimacy of the use of unmanned weapon systems for targeting. It deals with international legal, ethical and operational norms, as well

as the manner in which military institutions have incorporated them into their procedures. Mark Roorda joined the Royal Netherlands Marine Corps in 2004. He has obtained a Bachelor's degree in Military Sciences at the Netherlands Defence Academy and a Master's degree in Public International Law at Utrecht University. During operations in Afghanistan he gained experience in the targeting process in working with unmanned systems, and in applying legal norms in practice in his capacity as a forward air controller and artillery observer. He was also a participant in the Multinational Capability Development Campaign autonomous systems project.

Pertti Saariluoma is a professor of cognitive science in University of Jyväskylä, Finland. He holds a Ph.D. in psychology from the University of Turku (1984). He has studied and visited with universities in Oxford, Carnegie-Mellon, Cambridge, Aberdeen, Granada and Eindhoven, and the International Institute of Applied Systems Analysis in Austria. He was the first professor of cognitive science in Finland in the University of Helsinki and since 2001 he has been professor of cognitive science in University of Jyväskylä. His research interests concern expertise, thinking and intellectual processes, human-technology interaction and design thinking. He is member of Finland's Scientific Advisory Board for Defence.

Paul Scharre is a senior fellow and director of the 20YY Warfare Initiative at the Center for a New American Security. From 2008–13, Scharre worked in the Office of the Secretary of Defense (OSD), where he played a leading role in establishing policies on unmanned and autonomous systems and emerging weapons technologies. He led the Department of Defense (DoD) working group that drafted DoD Directive 3000.09, which established the department's policies on autonomy in weapon systems. He also led DoD efforts to establish policies on intelligence, surveillance, and reconnaissance (ISR) programmes and directed energy technologies. Scharre was involved in drafting policy guidance in the 2012 Defense Strategic Guidance, 2010 Quadrennial Defense Review, and secretary-level planning guidance. His most recent position was

special assistant to the under-secretary of defense for policy. Prior to joining OSD, Scharre served as a special operations reconnaissance team leader in the Army's 3rd Ranger Battalion and completed multiple tours to Iraq and Afghanistan. He is a graduate of the Army's Airborne, Ranger, and Sniper Schools and Honor Graduate of the 75th Ranger Regiment's Ranger Indoctrination Program. Scharre has published articles in *Foreign Policy*, *Politico*, *The National Interest*, *Proceedings*, *Armed Forces Journal*, *Joint Force Quarterly*, *Military Review*, and in academic technical journals. He has presented at the United Nations, NATO Defence College, Chatham House, National Defense University and numerous defence conferences on robotics and autonomous systems, lethal autonomous weapons, defence institution building, ISR, hybrid warfare, and the Iraq war. Scharre holds an M.A. in Political Economy and Public Policy and a B.S. in Physics, cum laude, both from Washington University in St. Louis.

Brandon Suarez currently serves as an engineer for the Aircraft Systems Group of General Atomics Aeronautical Systems, Inc. (GA-ASI), where he leads the company's Sense and Avoid (SAA) efforts, including system development, aircraft integration, and flight testing. The Aircraft Systems Group manufactures, supports, and operates a variety of proven, reliable, remotely piloted aircraft systems for military and commercial applications worldwide. In his current role, Suarez leads a collaborative team of experts from the US Federal Aviation Administration, National Aeronautics and Space Administration, and several industry partners to bring together the technology needed for a proof-of-concept SAA system on a Predator B unmanned aircraft system. He also is involved in the development of national and international SAA standards through the Radio Technical Commission for Aeronautics Special Committee (SC-228) and other community forums. The goal of these efforts is to write the standards that will be used to certify the unique features of unmanned air systems for flight in the US National Airspace System. Prior to joining GA-ASI in 2011, Suarez worked for Aurora Flight Sciences developing unconventional man-packable unmanned aircraft systems, some of which are being tested now in combat. He

holds Bachelor's and Master's degrees in Aerospace Engineering from the Massachusetts Institute of Technology.

Christoph Sulzbachner received his M.Sc. degrees in 2006 and 2008. The focus of his university education was on hardware and software development for real-time systems and information management in general. He joined the Austrian Institute of Technology (AIT) in 2008. His research interests include sensor technologies for autonomous aerial systems. Since 2012 he has been certified as a project management professional. Since 2013, he has served as vice-chair of research and development in the national UAV working group. He is the author and co-author of many scientific publications. He has experience in medium to large-scale projects, as he is working on and managing national and European-funded projects.

Erik Theunissen is a professor at the Netherlands Defence Academy. He has been active in the area of electronic navigation systems since 1991. He holds an M.Sc. in Aerospace Engineering and an M.Sc. and Ph.D. in Electrical Engineering. He is the author of over 130 papers on the topics of airspace integration, enhanced and synthetic vision systems, airport surface navigation, display design, and unmanned systems, and the recipient of more than 20 best paper awards. Prototypes of synthetic vision systems designed by his research team have been successfully evaluated in test aircraft such as the NASA ARIES, the Boeing Technology Demonstrator, the US Air Force Speckled Trout, and aircraft from the Federal Aviation Administration and Ohio University.

Andreas Tolk is chief scientist for SimIS Inc., a company based in Portsmouth, VA, United States. He is adjunct professor for Modelling, Simulation, and Visualization Engineering as well as for Engineering Management and Systems Engineering at Old Dominion University, Norfolk, VA. He holds a Ph.D. and M.S. in Computer Science, both from the University of the Federal Armed Forces of Germany in Munich. He received the first Technical Merit Award from the Simulation Interoperability Standards Organization in 2010, the Outstanding Professional Contributions

Award from the Society for Modelling and Simulation in 2012, followed by the Distinguished Professional Achievement Award in 2014. He has served on several NATO Modelling and Simulation Group and Systems Analysis and Studies activities, including being the technical evaluator for the annual NATO Modelling and Simulation Symposium between 2003 and 2011.

Andrew Williams is a section head in the Operational Analysis Branch in NATO's Headquarters Supreme Allied Commander Transformation. He has conducted numerous studies on a wide variety of defence subjects, including: policy impacts of autonomous and unmanned systems, military planning, operations assessment and programme evaluation, civil military interaction, private military security companies, organisational collaboration, and concept development and experimentation techniques. Prior to working at NATO, he served as an operational analyst for the UK Defence Science and Technology Laboratory. He holds a Master's of Physics from the University of Manchester, and a Ph.D. in Public Policy and Administration from Old Dominion University in Norfolk, VA. He has published articles in the *American Journal of Evaluation* and *Conflict and Cooperation*, amongst others, and was the lead editor of Volume 1 in the *Innovation in Capability Development* book series.

Christian Zinner is a researcher and thematic coordinator for the 3D Vision and Modelling research field at the Austrian Institute of Technology (AIT), where he evolves research and contract projects related to visual 3D object reconstruction, obstacle detection for autonomous trains, and sensor systems for autonomous trucks. His R&D activities go back to 1998, when he developed a finite element method software suite for the numerical analysis of structural mechanics and acoustics. In 2005 he joined the Safe and Autonomous Systems business unit at the AIT, where he investigates systematic performance optimization methods for image processing and computer vision in embedded real-time platforms.

PART 1:

INTRODUCTION

The Opportunity and Challenge of Autonomous Systems

Paul D. Scharre

Abstract

Autonomous systems pose both an opportunity and a challenge for warfighters. Militaries who harness their advantages will be able to field forces with greater range, persistence, mass, coordination, and speed on the battlefield. At the same time, autonomous systems, in their current state, lack the robustness and flexibility of human intelligence and are not appropriate for many tasks. Human-machine teaming to combine the best of each is likely to result in the most capable future force. But determining the appropriate balance between human and machine cognition will not always be easy. Autonomous systems are already capable enough to eliminate many tasks currently performed by warfighters, and are likely to continue improving over time. At the same time, war is and will remain a human endeavour. Robots are not combatants, but are merely tools in the hands of warfighters to better perform their job: to fight and win.

Introduction

Whether in cars, airplanes, stock-trading algorithms, or household appliances, machines across a range of industries are incorporating increasing intelligence and autonomy. Automated algorithms set book prices online.⁽¹⁾ Subway repairs are scheduled by tailored artificial intelligences.⁽²⁾ ‘Learning’ thermostats monitor your behaviour and adjust household temperature accordingly.⁽³⁾ And an estimated 61% of internet traffic is generated by automated and intelligent software called ‘bots’, from so-called Twitter bots that impersonate

1. Eisen 2011.

2. Hodson 2014.

3. Nest ND.

human users⁽⁴⁾ to ‘malware’ that maliciously attacks and disrupts systems or harvests information.⁽⁵⁾

Increasingly intelligent and autonomous systems have the potential to play a key role in military operations. More intelligent robotic technologies and algorithms will allow militaries to flood the battlespace with networks of cooperative, autonomous systems to overwhelm enemy defences. Software-based autonomous systems will help warfighters sift through massive amounts of data, shortening decision cycles and accelerating responses to enemy movements. And autonomous systems can perform tasks like taking off and landing from aircraft carriers with greater precision and reliability than humans, which can save money and warfighter lives.

These trends of increasing autonomy and intelligence in systems present both an opportunity and a challenge for warfighters and policy makers. On the one hand, such systems have the potential to deliver more effective and potentially more humane military force on the battlefield. Autonomy allows uninhabited systems to operate in environments where communications are degraded or denied. Robotic systems can operate untethered from the limits of human endurance, giving militaries that harness their advantages greater range and persistence on the battlefield – key attributes for countering the growing threat of long-range, precision-strike weapons. Commanders can also send robotic systems on more hazardous missions, for which they would not be willing to risk a human life, allowing entirely new concepts of operation. As uninhabited vehicles incorporate greater autonomy, militaries will be able to shift from today’s remote-controlled paradigm in which one person controls one vehicle, to a swarm paradigm, in which one person controls many vehicles at the mission level. This will allow militaries to field greater mass on the battlefield with the same personnel end strength. And increasing automation will allow military forces to operate with greater coordination, intelligence, and speed by shortening decision cycles or, in some cases, removing humans from them entirely.

On the other hand, however, increasing autonomy and intelligence in military systems also presents a challenge. Commanders will want to harness

4. Freitas et al. 2014; Isaacson 2011; Urbina 2013; McMillan 2012; Baraniuk 2014.

5. Zeigman 2013.

the advantages of autonomous systems without losing control over important military decisions. Determining which tasks or decisions should be delegated to autonomous systems and which should be reserved for humans may not be straightforward, however. Humans and machines excel at different cognitive tasks, and the best model will invariably be a blend of the two. As machine intelligence increases, however, the line between human and machine decision making will shift over time.

In the face of these challenges, a ‘go slow’ approach might be tempting. Experienced warfighters know that in combat, a less-capable weapon that works 100% of the time is preferred over a new, but unreliable, gadget. Autonomy, in particular, faces additional hurdles to acceptance because the very essence of automation is delegating tasks that were once done by people to machines. This requires trust – which is not easily given, and is rapidly taken away if (or when) autonomous systems fail. And inevitably, like all new technologies, autonomous systems often do not work well at first, and require iterative development between engineers and warfighters to determine the best system.

But there is an urgency to moving quickly. Unlike previous game-changing technologies like stealth or precision-guided weapons, which came from secret military labs, much of the innovation in autonomous systems is being driven by the commercial sector. Experts estimate that by 2018, global spending on military robotics will reach \$7.5 billion a year. In the same timeframe, global spending on commercial and industrial robotics is expected to top \$43 billion a year.⁽⁶⁾ Furthermore, much of the technology that enables autonomy – improved sensors, faster processing, lower-cost inertial measurement units, and miniaturization, among others – is driven by a broader revolution in information technology. Global investments in information technology totalled nearly \$3.8 trillion in 2014 – more than double total global military spending, including on personnel. Moreover, information technology spending is growing at a steady rate, while global military expenditures are flat or declining.⁽⁷⁾ As the information revolution continues, innovations will continue to have ‘spin-in’ opportunities to the defence sector, enabling better sensors, communications, and increasingly intelligent hardware and software.

6. Horowitz 2014.

7. For global defense expenditures, see SIPRI ND. For global information technology spending, see Gartner ND.

The result is a technology space that is globalized and accelerating. New innovations are widely available, to friend and foe alike, and the pace of innovation far outstrips that of most government bureaucracies, particularly when new weapons are often developed and acquired on decades-long timelines. Moreover, because the technology behind increasing autonomy is software rather than hardware based, it can be easily copied, modified, and proliferated. In fact, the best algorithms for controlling military systems may come from the commercial sector: spin-in technologies from intelligent video surveillance systems, driverless cars, or co-operative ‘swarming’ warehouse robots.⁽⁸⁾ For example, hobbyist drones available online for under \$1,000 have more autonomy than an MQ-9 Reaper military aircraft. While militaries are working to incorporate ‘remotely piloted’ aircraft into their forces, the paradigm of physically flying uninhabited aircraft remotely is already behind the curve technologically. Civilian hobbyists operate cheap drones that are able to take off, land, fly pre-programmed routes, and follow moving GPS-designated objects fully autonomously.⁽⁹⁾ Many companies are ready to field cooperative, autonomous systems that allow one person to control many vehicles at the same time.⁽¹⁰⁾ Military bureaucracies that build complex, intricate weapon systems on 20- or 30-year timelines, often with proprietary technologies, are at an inherent disadvantage in keeping pace with rapidly advancing information technology.⁽¹¹⁾

In a world where some of the most game-changing technologies will be widely available, uncovering the best uses of that technology – and doing so urgently – will be vital to sustaining NATO’s military dominance. Defence capability development and procurement processes need significant improvement in order to allow a more rapid refresh of information technology. Shorter acquisition cycles and modular design will be key components of a more agile

8. Ackerman 2012; Intelligent Security Systems ND; Seagate ND; Siemens ND.

9. For example, the fully autonomous 3DR Iris+ with ‘3PV Follow Me’ technology retails online for \$750. See <http://3drobotics.com/iris/> and <http://3drobotics.com/follow-me-info/>, accessed 23 March 2015.

10. For example, see aurora.aero ND; proxytechnologiesinc.com ND.

11. General Mike Hostage, commander of the Air Force’s Air Combat Command, has made this point about the software on the F-22, saying ‘The F-22, when it was produced, was flying with computers that were already so out of date you would not find them in a kid’s game console in somebody’s home gaming system. But I was forced to use that because that was the [specification] that was written by the acquisition process when I was going to buy the F-22.’ *Defense News* 2014. See also DARPA ND.

technology strategy. But understanding the best uses of autonomous systems will ultimately be what separates militaries that successfully capitalize on the advantage of autonomous systems from those that do not.

From the phalanx to the blitzkrieg, the history of disruptive innovations in warfare suggests that understanding how best to use a new technology is more important than developing the technology first, or even having the best technology. This places a premium on experimentation, prototyping, and an iterative process of concept development and technology improvement. Most importantly, it stresses the central importance of fostering a military culture that is willing to experiment with and embrace new ways of fighting. It is the potential to use emerging technologies in new ways that leads to truly revolutionary developments in warfare, not merely new widgets employed using old doctrine. When autonomous systems challenge existing doctrine or concepts of operation, the opportunities they present should be embraced.

Autonomous systems will not replace warfighters, any more than previous innovations like firearms, steam-powered ships, or tanks replaced combatants. These innovations did, however, change how militaries fight, and autonomous systems will too. Autonomous systems will inevitably change how some military duties are performed, and in some cases may eliminate some job specialties entirely. Physical prowess for some tasks – like piloting an aircraft, driving a vehicle, or firing a rifle – will be less important in a world where aircraft fly themselves, vehicles drive on their own, and smart rifles correct for wind, humidity, elevation, and the shooter's movements all on their own.⁽¹²⁾ The ethos embodied in these job specialties will not change, however, and human judgment will always be required in combat. Today's infantrymen, sailors, and cavalymen no longer fight with edged weapons, work the sails and rigging of ships, or ride horses, but the ethos embodied in their job specialties lives on, even as the specific ways in which warfighters carry out those duties have changed. Similarly, the duties of tomorrow's 'pilots', 'tank drivers', and 'snipers' will look far different from today, but the need for human judgment and decision making in combat will not change.

Understanding how best to incorporate autonomous systems into military forces requires cross-disciplinary collaboration between not only the military

12. All of which exist today. See <http://tracking-point.com/>, accessed 23 March 2015.

practitioners and engineers who design the autonomous systems, but also lawyers, ethicists, and policy makers when the use of autonomy raises legal, ethical, or policy concerns. Many questions surrounding the incorporation of autonomous systems into military operations centre not on whether autonomous systems *can* perform a task, but whether they *should*. As such, this volume incorporates a wide range of views, some of which offer competing perspectives on the appropriate role of autonomous systems in future military operations. As a preface to these perspectives, this introduction will briefly introduce three core concepts: the definition of autonomy, autonomy in the use of force, and human–machine teaming.

What is Autonomy?

What is the nature of autonomy? What does it mean for a robot, or any military system, to be ‘fully autonomous’? How much machine intelligence is required to reach ‘full autonomy’, and when can we expect it? And what is the role of the human warfighter with respect to autonomous systems?

Confusion about the term ‘autonomy’ itself is a hindrance to envisioning answers to these questions. People use the word autonomy in different ways, which makes communicating about where militaries are headed with autonomous systems particularly challenging. The term ‘autonomous robot’, for example, might mean a house cleaning Roomba robot to one person and a science fiction Terminator to another! Writers or presenters on this topic often articulate ‘levels of autonomy’, but their levels rarely agree, leading a recent US Defense Science Board report on autonomy to recommend rejecting the concept of ‘levels’ of autonomy altogether.⁽¹³⁾

In its simplest form, autonomy is the ability of a machine to perform a task without human input. Thus an ‘autonomous system’ is a machine, whether hardware or software, that, once activated, performs some task or function on its own. A robot combines an uninhabited platform or vehicle with some degree of autonomy, which is generally understood to include the ability to sense and react to the environment, at least in some crude fashion.

Autonomous systems are not limited to robots or uninhabited vehicles, however. In fact, autonomous, or automated, functions are included in many

13. Defense Science Board 2012.

human-inhabited systems today. Most cars today include anti-lock brakes, traction and stability control, power steering, emergency seat belt retractors, and air bags. Higher-end cars may also include intelligent cruise control, automatic lane keeping, collision avoidance, and automatic parking. For military aircraft, automatic ground collision avoidance systems (auto-GCAS) can similarly take control of a human-piloted aircraft if a pilot becomes disoriented and is about to fly into terrain.⁽¹⁴⁾ And modern commercial airliners have a high degree of automation available throughout every phase of a flight. Increased autonomy can have many advantages, including:

- increased safety and reliability;
- improved reaction time and performance;
- reduced personnel burden, with operational advantages or cost savings; and
- the ability to continue operations in communications-degraded or -denied environments.

Parsing out how much autonomy a system has is important for understanding the challenges and opportunities associated with increasing autonomy. There is a wide gap, of course, between a Roomba and a Terminator. Rather than search in vain for a unified framework of ‘levels of autonomy’, a more fruitful direction is to think of autonomy as having three main axes, or dimensions, along which a system can vary. These dimensions are independent, and so autonomy does not vary along a single spectrum, but rather along all three spectrums simultaneously. What makes understanding autonomy so difficult is that people tend to use the same word to refer to three completely different concepts:

1. the human-machine command-and-control relationship;
2. the sophistication of the machine’s decision making; and
3. the type of decision or function being automated.

14. Auto-GCAS had its first “save” in February 2015 over Syria. See Norris 2015.

The Human–Machine Command-and-Control Relationship

The first dimension along which an autonomous system can vary is the relationship between the human and the machine. Machines that perform a function for some period of time, then stop and wait for human input before continuing, are often referred to as ‘semiautonomous’ or ‘human *in* the loop’. Machines that can perform a function entirely on their own but have a human in a monitoring role, who can intervene if the machine fails or malfunctions, are often referred to as ‘human-supervised autonomous’ or ‘human *on* the loop’. Machines that can perform a function entirely on their own and humans are unable to intervene are often referred to as ‘fully autonomous’ or ‘human *out* of the loop’.

The Sophistication of the Machine’s Decision Making

The word ‘autonomy’ is also used in a completely different way to refer to the sophistication of a system’s decision making. Regardless of the human–machine command-and-control relationship, words such as ‘automatic’, ‘automated’, and ‘autonomous’ are often used to refer to the spectrum of complexity of machines. The term ‘automatic’ is often used to refer to systems that have very simple, mechanical responses to environmental input, such as trip wires, mines, toasters, and old mechanical thermostats. The term ‘automated’ is often used to refer to more complex, rule-based systems, for example self-driving cars and modern programmable thermostats. Sometimes the word ‘autonomous’ is reserved for machines that execute some kind of self-direction, self-learning, or emergent behaviour that was not directly predictable from an inspection of its code. An example would be a self-learning robot that taught itself how to walk or the Nest ‘learning thermostat’.⁽¹⁵⁾

Others reserve the word ‘autonomous’ only for entities that have intelligence and free will, but these concepts hardly add clarity. Artificial intelligence is a loaded term that can refer to a wide range of systems, from those that exhibit near-human or superhuman intelligence in a narrow domain – such as playing chess (Deep Blue), playing Jeopardy (Watson), or programming subway repair schedules – to potential future systems that might have human or superhuman

15. Lipson 2007; Nest ND.

general intelligence.⁽¹⁶⁾ But whether general intelligence leads to free will, or whether humans even have free will, is itself debated.

What is particularly challenging is that there are no clear boundaries between these degrees of complexity, from ‘automatic’ to ‘automated’ to ‘autonomous’ to ‘intelligent’, and different people may disagree on what to call any given system.

Type of Decision or Function being Automated

Ultimately, it is meaningless to refer to a machine as ‘autonomous’ or ‘semiautonomous’ without specifying the task or function being automated. Different decisions have different levels of complexity and risk. A mine and a toaster have radically different levels of risk, even though both have humans ‘out of the loop’ once activated and both use very simple mechanical switches. The task being automated, however, is much different. Any given machine might have humans in complete control of some tasks and might autonomously perform others. For example, an ‘autonomous car’ drives from point A to point B by itself, but a person still chooses the final destination. The car is only autonomous with respect to some functions.

From this perspective, the question of when we will get to ‘full autonomy’ is meaningless. There is not a single spectrum along which autonomy moves. A better framework would be to ask which tasks are done by a person, and which by a machine, and what is the relationship of the person to the machine for various tasks. Thus, it is more fruitful to think about ‘autonomous functions’ of systems, rather than characterizing an entire vehicle or system as ‘autonomous.’⁽¹⁷⁾

Importantly, these three dimensions of autonomy are independent. The intelligence or complexity of the machine is a separate concept from the tasks it performs. Increased intelligence or more sophisticated machine reasoning to perform a task does not necessarily equate to transferring control over more tasks from the human to the machine. Similarly, the human–machine command-and-control relationship is a different issue from complexity or tasks performed. A thermostat functions on its own without any human supervision

16. On Deep Blue, see IBM, ‘Icons of Progress: Deep Blue’, ND. On Watson, see IBM, ‘Icons of Progress: A Computer Called Watson’, ND. Also see Hodson 2014; Barrat 2013.

17. MCDC 2014.

or intervention when you leave your house, but it still has a limited set of functions it can perform.

Autonomy, in this respect, is better thought of as a general attribute of a system, like propulsion or communications, rather than a defining feature of a system. Just as there are different types of propulsion and communications, each with different strengths and advantages appropriate for different applications, there are many different ways in which autonomy can be employed in military systems. Thus the current prevalence of the term ‘autonomous system’ reflects the current interest in autonomy, rather than a necessary description of the system. It would be strange to refer to an aircraft as a ‘communications-enabled’ platform.

Autonomy in the Use of Force

The use of autonomy in functions relating to the use of force raises particularly challenging issues. Autonomy is currently used in a range of engagement-related functions, including: acquiring, tracking, identifying, and cueing potential targets; aiming weapons; prioritizing targets to be engaged; timing of when to fire or release weapons; manoeuvring and homing in on targets; and detonation. A pilot conducting a beyond-visual-range air-to-air engagement, for example, leverages automation in sensing and identifying the enemy aircraft, as well as in assisting the missile in homing in on the enemy aircraft once the pilot has made the decision to engage. Generally speaking, however, humans remain in control of decisions about whether to use force as well as the specific targets selected for engagement.⁽¹⁸⁾ Weapons that select and engage targets on their own, without a human decision to engage those specific targets, would be a significant shift. The US Department of Defense, the International Committee of the Red Cross (ICRC), and the United Nations (UN) Special Rapporteur for

18. There have been some exceptions to this general pattern, specifically a handful of loitering wide-area munitions that select and engage specific targets on their own. These include the U.S. Tomahawk Anti-Ship Missile, a wide-area loitering anti-radar weapon designed to target Soviet ships during Cold War. It was retired from the U.S. inventory in the 1990s. Experimental, but not deployed, systems include the U.S. Tacit Rainbow loitering anti-radar weapon and Low-Cost Autonomous Attack System (LOCAAS), designed to find and destroy enemy tanks. The Israeli Harpy is a loitering anti-radar weapon that searches over a wide area for enemy radars and then, once it finds one meeting its target parameters, flies into it, destroying the radar. It has been sold to Turkey, South Korea, China, and India, and the Chinese are reported to have reverse-engineered their own variant. See Kopp 2012; Parsch 2006; National Museum of the U.S. Air Force 2011; Israel Aerospace Industries ND; *Defense Update* 2006; Chinese Military Aviation Blog 2014.

extrajudicial, summary or arbitrary executions refer to weapons of this type as ‘autonomous weapon systems’, or simply ‘autonomous weapons’.⁽¹⁹⁾

Autonomous Weapons Raise Challenging Issues

Autonomous weapons that would select and engage targets on their own raise challenging legal, moral, ethical, and policy considerations. The prospect of potential future autonomous weapons has sparked a UN special rapporteur report, a US Department of Defense policy directive, an International Committee for Robot Arms Control, a consortium of over 50 non-governmental organizations as part of a ‘Campaign to Stop Killer Robots’, engagement from the ICRC, and a host of articles and reports from lawyers, ethicists, roboticists, military officers, and defence policy experts.⁽²⁰⁾ In the spring of 2014, state parties to the UN Convention on Certain Conventional Weapons (CCW), which has previously banned or regulated weapons such as blinding lasers and incendiary weapons, held the first multilateral discussions on autonomous weapons.⁽²¹⁾ Another round of discussions at the CCW was held in April 2015.

From a legal perspective, autonomous weapons, like all weapons, must be used in a manner compliant with the laws of armed conflict (LOAC), which includes meeting the principles of distinction, proportionality, precautions in attack, avoiding superfluous injury, and respecting *hors de combat*. An autonomous weapon that was unable to comply with these principles – say, if it could not distinguish enemy combatants from civilians – would not be lawful for use in an environment in which that function was anticipated to be needed. But there are no prohibitions against autonomous weapons, *per se*. If an autonomous weapon could be used in a manner consistent with existing LOAC principles, then it would be lawful for use in that situation.⁽²²⁾

19. DOD 2012; Heyns 2013; ICRC 2014.

20. The official page for the Campaign to Stop Killer Robots is available at: <http://www.stopkillerrobots.org/>, accessed 24 March 2015. The official page of the International Committee for Robot Arms Control is <http://icrac.net/>, accessed 24 March 2015. For a short primer on the case against autonomous weapons by activists, see HRW 2012. For a comprehensive bibliography on autonomous weapons, see CNAS ND.

21. United Nations Office at Geneva, ‘The Convention on Certain Conventional Weapons’, ND; United Nations Office at Geneva, ‘CCW – Latest Information’, ND.

22. For an excellent overview of the legal issues surrounding autonomous weapons, see Anderson, Reisner, and Waxman 2014; Anderson and Waxman 2013.

Autonomous weapons do raise other issues, however. If it malfunctioned, who would be responsible? Would their use lead to an off-loading of moral responsibility for killing, and therefore more use? And can they be used in a manner that is safe and avoids unacceptable operational risk, including the risk of fratricide or susceptibility to cyber attack?

The Accelerating Pace of Battle

Simplistic conclusions about the appropriate role of autonomy and human control in the use of force are belied by the fact that many militaries already employ defensive, human-supervised autonomous weapon systems to defend vehicles, ships, and installations from short-warning attacks. These include air and missile defence systems; counter-rocket, artillery and mortar systems; and active-protection systems for ground vehicles. All told, at least 30 nations have such systems deployed or in development.⁽²³⁾ They are essential for responding to short-warning saturation attacks from missiles or rockets, where the time of engagement is too short for humans to adequately respond. They have been used narrowly to date, however: only for immediate defence of human-occupied bases or vehicles, and where human controllers can supervise their operation in real time and terminate engagements if necessary. Moreover, because human controllers are physically co-located with the system, either on a vehicle or a land base, they have physical access and can exercise hardware-level over-rides in the event of a software malfunction or cyber attack.

These defensive systems point to the pressures that might drive militaries to autonomous weapons: compressed engagement timelines and the need to defend against short-warning saturation attacks. Automated decision making may not always be as good as human decision making, but it need not be if it is faster, and if that speed leads to a sufficient advantage on the battlefield. Potential offensive uses for autonomous weapons include finding and destroying mobile targets with loitering munitions or uninhabited vehicles in communications-denied environments. And some potential uses defy easy characterization into 'offensive' or 'defensive' roles, such as uninhabited vehicles operating forward in communications-denied environments that use force to defend themselves from attack, firing back if fired upon or, conceivably, shooting first.

23. Scharre and Horowitz 2015.

The question of how much autonomy to incorporate in weapons is particularly challenging because it occurs in a ruthlessly competitive environment. Increased autonomy that shortens decision cycles is a boon when employed on behalf of one's own forces, but is a threat when used by the enemy. When employed by both sides in a conflict, ever-increasing automation raises the prospect of an accelerated pace of battle, with humans struggling to keep pace with the faster reaction times of machines.

'Flash Wars' and Fragile Stability

The prospect of a quickening pace of battle that threatens to take control increasingly out of the hands of humans raises serious concerns. Just as the unanticipated interaction of automated trading algorithms has led to stock market 'flash crashes', the interaction of autonomous systems in military crises could introduce instabilities. The lure of quicker reaction times, or merely the fear that other nations might develop autonomous weapon systems, could spark an autonomous weapons arms race. This potential 'gunslinger' quality of autonomy is exceptionally dangerous and destabilizing, particularly in cyberspace, where operations move at 'net speed'.

There is a tension between the speed of operations and the speed of human decision making. While militaries will need to embrace autonomy for some purposes, humans must also be kept in the loop for the most critical decisions, particularly those that involve the use of force or movements and actions that could potentially be escalatory in a crisis. Autonomy that might make sense tactically would be disastrous strategically if it led to 'flash wars'.

During the Cold War, defence planners faced a similar problem of 'fragile stability', whereby vulnerable nuclear arsenals incentivized an enemy to strike first. In response, military strategists developed a doctrine of an assured second-strike capability in order to reduce the incentives for a first strike. Similarly, strategists today must focus on resiliency in order to be able to absorb a sudden destabilizing attack and buy time for decision makers to understand a crisis before deciding how to respond. While autonomy will be essential for some purposes, it should not take the place of human decisions about when and how to use force.

Autonomous Weapons and the Military Profession

While ‘flash wars’ driven by an arms race in autonomy may seem a distant and unlikely possibility, it is inevitable that, as militaries incorporate greater autonomy into a wide range of systems, they will be forced to grapple with the question of autonomous weapons. At least 30 nations⁽²⁴⁾ today either have or are developing armed uninhabited aircraft, and such technology is within the reach of any modern military. When communications links are severed, what will those aircraft be programmed to do? Return to base? Perform non-lethal missions like reconnaissance and surveillance? Strike pre-authorized fixed targets, much like cruise missiles today? Hunt and destroy mobile targets of opportunity? And if threatened or fired upon, will they be programmed to fire back?

Militaries will weigh these decisions within the broader milieu of lawyers, activists, ethicists, and public opinion, but ultimately some of the most important issues regarding autonomous weapons are ones of operational risk and military ethics, and are therefore the province of defence professionals. Unlike other forms of autonomy, such as self-driving vehicles or autonomous aircraft, increased autonomy in the use of force goes right to the heart of the essence of the military profession: expertise in decisions about the application of force in war. Autonomous weapons change the relationship of the military professional to the use of force, ceding control over the decision to engage specific targets with force from the warfighter on the battlefield to engineers, testers, and evaluators who designed and pre-programmed target selection specifications into the weapons. Assuming that such decisions can be automated safely and effectively, whether they *should* be is not principally a legal matter, but rather one of command responsibility and military professional ethics.

Human–Machine Teaming

Autonomous systems will not replace warfighters, but they can help warfighters better perform their missions and accomplish their military objectives. Autonomous systems will be able to perform many military tasks better than humans, and will be particularly useful in situations where speed and precision are required, or where repetitive tasks are to be performed in relatively structured

24. Horowitz and Fuhrmann 2014

environments. At the same time, barring major advances in novel computing methods that aim to develop computers that work like human brains, such as neural networks or neuromorphic computing, autonomous systems will have significant limitations. While machines exceed human cognitive capacities in some areas, particularly speed, they lack robust general intelligence that is flexible across a range of situations. Machine intelligence is ‘brittle’. That is, autonomous systems can often outperform humans in narrow tasks, such as chess or driving, but if pushed outside their programmed parameters they fail, and often badly. Human intelligence, on the other hand, is very robust to changes in the environment and is capable of adapting and handling ambiguity. As a result, some decisions, particularly those requiring judgment or creativity, will be inappropriate for autonomous systems. The best cognitive systems, therefore, are neither human nor machine alone, but rather human and machine intelligences working together.

Militaries looking to best harness the advantages of autonomous systems should take a cue from the field of ‘advanced chess’, where human and machine players cooperate in hybrid or ‘centaur’ teams. After world chess champion Gary Kasparov lost to IBM’s chess-playing computer Deep Blue in 1996 (and again in a 1997 rematch), he founded the field of advanced chess, which is now the cutting edge of chess competition. In advanced chess, humans play in cooperation with a computer chess program on a team. The humans can use the program to evaluate possible moves and try out alternative sequences. The result is a superior game of chess, more sophisticated than would be possible with either humans or machines playing alone.⁽²⁵⁾

Human–machine teaming raises new challenges, and militaries will need to experiment to find the optimum mix of human and machine cognition. Determining which tasks should be done by machines and which by people will be an important consideration, and one made continually challenging by the fact

25. Humans can still add value to a combined human-computer chess team, although this may not be the case forever. Checkers is an entirely ‘solved’ game, so computers can play checkers perfectly and humans add no value. Chess is a much more complicated game than checkers, but machine abilities at chess continue to grow. For an interesting explanation of the advantages of human and machine cognition in chess today, see Cowen 2013.

that machines continue to advance in cognitive abilities.⁽²⁶⁾ Human-machine interfaces and training for human operators to understand autonomous systems will be equally important. Human operators will need to know the strengths and limitations of autonomous systems, and in which situations autonomous systems are likely to lead to superior results and when they are likely to fail. As autonomous systems become incorporated into military forces, the tasks required of humans will change, not only with respect to what functions they will no longer perform, but also which new tasks will be demanded of them. Human operators will need to be able to understand, supervise, and control complex autonomous systems in combat.⁽²⁷⁾ This places new burdens on the selection, training, and education of military personnel, and potentially raises additional policy concerns. Cognitive human performance enhancement may help and may in fact be essential to managing the data overload and increased operations tempo of future warfare, but has its own set of legal, ethical, policy, and social challenges.

How militaries incorporate autonomous systems into their forces will be shaped in part by strategic need and available technology, but also in large part by military bureaucracy and culture. Humans may be unwilling to cede control over some tasks to machines. Debates over autonomous cars are an instructive example. Human beings are horrible drivers, killing more than 30,000 people a year in the United States alone, or roughly the equivalent of a 9/11 attack every month. Self-driving cars, on the other hand, have already driven nearly three-quarters of a million miles, including in crowded city streets, without a single accident.⁽²⁸⁾ Autonomous cars have the potential to save literally tens of thousands of lives every year, yet rather than rushing to put self-driving cars on the streets as quickly as possible, adoption is moving forward cautiously.⁽²⁹⁾ Given the state of the technology today, even if autonomous cars are far better than human drivers overall, there would inevitably be situations in which the

26. Humans are also increasing in intelligence, although not at the same rate. Even though the underlying genetic hardware of the brain has not changed significantly, human cognition has measurably increased, on the order of 30 IQ points, over the past 100 years. Whether this increase can be extrapolated into the future is unclear. See Flynn 2013.

27. For more on what this shift to supervisory control will mean and the challenges it will bring, see Hawley et al. 2005.

28. Urmson 2014.

29. *The Star* 2014.

autonomy fails and humans, who are better at adapting to novel and ambiguous circumstances, would have done better in that instance. Even if, in aggregate, thousands of lives could be saved with more autonomy, humans tend to focus on the few instances in which humans would have performed better.⁽³⁰⁾ Ceding human control to autonomous systems requires trust, which is not easily given.

The Human Element

A reluctance to embrace autonomous systems often stems from a perception that they are replacing humans, and terminology that refers to ‘unmanned’ systems can feed this perception. The reality, however, is a future of human–machine teaming. Many of the tasks that humans perform in warfare will change, but humans will remain central to war, for good or ill. The introduction of increasingly capable autonomous systems on the battlefield will not lead to bloodless wars of robots fighting robots, with humans sitting safely on the sidelines. Death and violence will remain an inescapable component of war, if for no other reason than that it will require real human costs for wars to come to an end. Nor will humans be removed from the battlefield entirely, telecommuting to combat from thousands of miles away. Remote operations will have a role, as they already do in uninhabited aircraft operations today, but humans will be needed forward in the battlespace, particularly for command and control when long-range communications are degraded.

Even as autonomous systems play an increasing role on the battlefield, it is still humans who will fight wars, only with different weapons. Combatants are people, not machines. Technology will help humans fight, as it has since the invention of the sling, the spear, and the bow and arrow. Better technology can give combatants an edge in terms of standoff, survivability, or lethality, advantages that combatants have sought since the first time a human picked up a club to extend his reach against an enemy. But technology alone is nothing without insight into the new uses it unlocks. The tank, radio, and airplane were critical components of the *blitzkrieg*, but the *blitzkrieg* also required doctrine, organization, concepts of operation, experimentation, and training to be developed successfully. And so it has been with all disruptive innovation on the battlefield. It is people who develop new concepts for fighting, draft

30. Lin 2013.

requirements for the technology, restructure organizations, rewrite doctrine, and ultimately fight. In the future, it will be no different.

War will remain a clash of wills. To the extent that autonomous systems allow more effective battlefield operations, they can be a major advantage. Those who master a new technology and its associated concepts of operation first can gain game-changing advantages on the battlefield, and achieve decisive victory over those who lag behind. But technological innovation in war can be a double-edged sword. If this advantage erodes a nation's willingness to squarely face the burden of war, it can be a detriment. The illusion that such advantages can lead to quick, easy wars can be seductive, and those who succumb to it may find their illusions shattered by the unpleasant and bloody realities of war.⁽³¹⁾ Autonomous systems can lead to advantages over one's enemy, but the millennia-long evolution of weapons and countermeasures suggests that such weapons will proliferate: no innovation leaves its user invulnerable for very long. In particular, increasing automation has the potential to accelerate the pace of warfare, but not necessarily in ways that are conducive to the cause of peace. An accelerated tempo of operations may lead to combat that is more chaotic, but not more controllable. Wars that start quickly may not end quickly.

Autonomous systems raise challenging operational, strategic, and policy issues, the full scope of which cannot yet be seen. The nations and militaries that see furthest into a dim and uncertain future to anticipate these challenges and prepare for them now will be best poised to succeed in the warfighting regime to come.

31. For an excellent critique of the belief that technology can make war quick or easy, see McMaster 2013.

References

- Ackerman, Evan. "Warehouse robots get smarter with ant intelligence". *IEEE Spectrum*, 26 March 2012. Available at <http://spectrum.ieee.org/automation/robotics/industrial-robots/warehouse-robots-get-smarter-with-ant-intelligence>, accessed 23 March 2015.
- Anderson, Kenneth, and Waxman, Matthew C. 'Law and Ethics for Autonomous Weapons: Why a Ban Won't Work and How the Laws of War Can', *Columbia University Public Law & Legal Theory Research Paper Series*, No. 13-351, April 2013. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2250126, accessed 23 March 2015.
- Anderson, Kenneth, Reisner, Daniel, and Waxman, Matthew C. 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems'. *International Law Studies* 90 (September 2014): 386 –411. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477095, accessed 23 March 2015.
- aurora.aero. "Aurora's Autonomy and Flight Control". Available at <http://www.aurora.aero/Research/Autonomy.aspx>, accessed 23 March 2015.
- Baraniuk, Chris. "How Online 'Chatbots' Are Already Tricking You". *BBC*, 9 June 2014. Available at <http://www.bbc.com/future/story/20140609-how-online-bots-are-tricking-you>, accessed 23 March 2015.
- Barrat, James. *Our Final Invention: Artificial Intelligence and the End of the Human Era*. New York: Thomas Dunne Books, 2013.
- Center for a New American Security (CNAS). 'Ethical Autonomy Bibliography', ND. Available at <http://www.cnas.org/research/defense-strategies-and-assessments/20YY-Warfare-Initiative/Ethical-Autonomy/bibliography>, accessed 24 March 2015.
- Chinese Military Aviation Blog*. 'UAV/UCAV', 2014. Available at <http://chinese-military-aviation.blogspot.ch/p/uav.html>, accessed 23 March 2015.
- Cowen, Tyler 'What Are Humans Still Good For? The Turning Point in Freestyle Chess May Be Approaching', *Marginal Revolution blog*, 5 November 2013. Available at <http://marginalrevolution.com/marginalrevolution/2013/11/what-are-humans-still-good-for-the-turning-point-in-freestyle-chess-may-be-approaching.html>, accessed 23 March 2015.

- DARPA, “Contrasted with Other Industries”. ND. Available at <http://i1.wp.com/blog.usni.org/wp-content/uploads/2013/08/DARPA-SLIDE.png>, accessed 23 March 2015.
- Defense News*. “Interview: Gen. Michael Hostage, Commander, US Air Force's Air Combat Command”. 3 February 2014. Available at <http://mobile.defense-news.com/article/302030017>, accessed 23 March 2015.
- Defense Science Board, *Task Force Report: The Role of Autonomy in DoD Systems*. Washington, DC: DOD, 2012. Available at <http://www.acq.osd.mil/dsb/reports/AutonomyReport.pdf>, accessed 23 March 2015.
- Defense Update*. ‘Harpy Air Defense Suppression System’, 4 March 2006. Available at <http://defense-update.com/directory/harpy.htm>, accessed 23 March 2015.
- Eisen, Michael. ‘Amazon’s \$23,698,655.93 Book about Flies’. *it is NOT junk*, 22 April 2011. Available at <http://www.michaeleisen.org/blog/?p=358>, accessed 23 March 2015.
- Flynn, James. ‘Why our IQ Levels are Higher than our Grandparents’. TED Talk, Long Beach, CA, 2013. Available at http://www.ted.com/talks/james_flynn_why_our_iq_levels_are_higher_than_our_grandparents, accessed 23 March 2015.
- Freitas, Carlos A., Ghosh, Saptarshi, Benevenuto, Fabricio, and Veloso, Adriano. “Reverse Engineering Socialbot Infiltration Strategies in Twitter”. 20 May 2014. Available at <http://arxiv.org/pdf/1405.4927v1.pdf>, accessed 23 March 2015.
- Gartner, “Gartner Worldwide IT Spending Forecast”. Available at <http://www.gartner.com/technology/research/it-spending-forecast/>, accessed 23 March 2015.
- Hawley, John et al. *The Human Side of Automation: Lessons for Air Defense Command and Control*. Army Research Laboratory, March 2005.
- Heyns, Christof. Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. Geneva and New York: United Nations General Assembly, 2013. Available at http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf, accessed 24 March 2015.

- Hodson, Hal. "The AI boss that Deploys Hong Kong's Subway Engineers," *New Scientist*, July 4, 2014. Available at <http://www.newscientist.com/article/mg22329764.000-the-ai-boss-that-deploys-hong-kongs-subway-engineers.html#.VRB7jELVt0c>, accessed 23 March 2015.
- Horowitz, Michael C. "The Looming Robotics Gap". *ForeignPolicy.com*, 5 May 2014. Available at http://www.foreignpolicy.com/articles/2014/05/05/the_looming_robotics_gap_us_military_technology_dominance, accessed 23 March 2015.
- Michael C. Horowitz and Matthew Fuhrmann, "Droning On: Explaining the Proliferation of Unmanned Aerial Vehicles," October 24, 2014.
- Human Rights Watch (HRW). *Losing Humanity: The Case Against Killer Robots*. New York: HRW, 2012. Available at http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf, accessed 24 March 2015.
- IBM, 'Icons of Progress: Deep Blue', ND. Available at <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/>, accessed 23 March 2015.
- IBM, 'Icons of Progress: A Computer Called Watson', ND. Available at <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/watson/>, accessed 23 March 2015.
- Intelligent Security Systems. Available at <http://isscctv.com>, accessed 23 March 2015.
- International Committee of the Red Cross (ICRC). *Report of the ICRC Expert Meeting on "Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects"*, 26-28 March 2014. Geneva: ICRC, 2014. Available at <https://www.icrc.org/eng/assets/files/2014/expert-meeting-autonomous-weapons-icrc-report-2014-05-09.pdf>, accessed 24 March 2015.
- Isaacson, Andy. "Are You Following a Bot?". *The Atlantic*, 2 April 2011. Available at <http://www.theatlantic.com/magazine/archive/2011/05/are-you-following-a-bot/308448/>, accessed 23 March 2015.
- Israel Aerospace Industries. 'Harpy Loitering Weapon', ND. Available at <http://www.iai.co.il/2013/16143-16153-en/IAI.aspx>, accessed 23 March 2015.
- Kopp, Carlo. 'Tomahawk Cruise Missile Variants: BGM/RGM/AGM-109 Tomahawk/TASM/TLAM/GCLM/MRASM: Technical Report

- APA-TR-2005-0702'. Air Power Australia, 2012. Available at <http://www.ausairpower.net/Tomahawk-Subtypes.html>, accessed 23 March 2015.
- Lin, Patrick. 'The Ethics of Autonomous Cars', *TheAtlantic.com*, 8 October 2013. Available at <http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/>, accessed 23 March 2015.
- Lipson, Hod. 'Building 'Self-aware' Robots', TED Talk, March 2007. Available at http://www.ted.com/talks/hod_lipson_builds_self_aware_robots, accessed 23 March 2015.
- McMaster, H.R. 'The Pipe Dream of Easy War', *New York Times*, 10 July 2013. Available at <http://www.nytimes.com/2013/07/21/opinion/sunday/the-pipe-dream-of-easy-war.html?pagewanted=all&r=0>, accessed 23 March 2015.
- McMillan, Robert. "Twitter Bots Fight It Out to See Who's the Most Human". *Wired*, 7 November 2012. Available at <http://www.wired.com/2012/11/twitter-bots/all/>, accessed 23 March 2015.
- Multinational Capability Development Campaign (MCDC). *Role of Autonomous Systems in Gaining Operational Access*. Norfolk, VA: Supreme Allied Commander Transformation HQ, 2014. Available at <http://innovationhub-act.org/sites/default/files/u4/Policy%20Guidance%20-%20Autonomy%20in%20Defence%20Systems%20MCDC%202013-2014.pdf>, accessed 23 March 2015.
- National Museum of the U.S. Air Force. 'Northrop AGM-136A Tacit Rainbow', 3 August 2011. Available at <http://www.nationalmuseum.af.mil/factsheets/factsheet.asp?id=418>, accessed 23 March 2015.
- Nest. "Life with Nest Thermostat". Available at <https://nest.com/thermostat/life-with-nest-thermostat/>, accessed 23 March 2015.
- Norris, Guy. "Ground Collision Avoidance System 'Saves' First F-16 In Syria," *Aviation Week*, 5 February 2015. Available at <http://aviationweek.com/defense/ground-collision-avoidance-system-saves-first-f-16-syria/>, accessed 18 April 2015.
- Parsch, Andreas. 'Lockheed Martin LOCAAS'. In *Directory of U.S. Military Rockets and Missiles*, 2006. Available at <http://www.designation-systems.net/dusrm/app4/locaas.html>, accessed 23 March 2015.

- proxytechnologiesinc.com. "Product and Services: Cooperative Control/UDMS". Available at <http://www.proxytechnologiesinc.com/systems.html>, accessed 23 March 2015.
- Seagate. "Security Gets Smarter With Intelligent Video Surveillance Systems". Available at <http://www.seagate.com/tech-insights/video-security-gets-smarter-with-intelligent-video-surveillance-systems-master-ti/>, accessed 23 March 2015.
- Scharre, Paul, and Horowitz, Michael C. *An Introduction to Autonomy in Weapon Systems*. Washington, DC: Center for a New American Security, 2015. Available at http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf, accessed 23 March 2015.
- Siemens, "Video Analytics Provide Intelligent Video Surveillance". Available at <http://www.buildingtechnologies.siemens.com/bt/global/en/security-solution/video-analytics/pages/video-analytics.aspx>, accessed 23 March 2015.
- Star, The. 'Self-driving Cars Take a Small Step Closer to Reality', 15 September 2014. Available at <http://www.thestar.com.my/Tech/Tech-News/2014/09/15/Self-driving-cars-take-a-small-step-closer-to-reality/>, accessed 23 March 2015.
- Stockholm International Peace Research Institute (SIPRI). "SIPRI Military Expenditures Database". Available at http://www.sipri.org/research/armaments/milex/milex_database, accessed 23 March 2015.
- United Nations Office at Geneva. 'The Convention on Certain Conventional Weapons', ND. Available at [http://www.unog.ch/80256EE600585943/\(httpPages\)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument), accessed 23 March 2015.
- United Nations Office at Geneva. 'CCW – Latest Information', ND. Available at <http://www.unog.ch/80256EE600585943/%28httpPages%29/3CFCEEEF-52D553D5C1257B0300473B77?OpenDocument>, accessed 23 March 2015.
- Urbina, Ian. "I Flirt and Tweet. Follow Me at #Socialbot". *New York Times*, 10 August 2013. Available at http://www.nytimes.com/2013/08/11/sunday-review/i-flirt-and-tweet-follow-me-at-socialbot.html?_r=0, accessed 23 March 2015.

Urmson, Chris. 'The Latest Chapter for the Self-Driving Car: Mastering City Street Driving', *Google Official Blog*, 28 April 2014. Available at <http://google-blog.blogspot.com/2014/04/the-latest-chapter-for-self-driving-car.html>, accessed 23 March 2015.

US Department of Defense (DOD). *Directive: Autonomy in Weapon Systems*, 3000.09, 21 November 2012. Available at <http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf>, accessed 24 March 2015.

Zeigman, Igal. "Report: Bot Traffic Is up to 61.5% of All Website Traffic". *Incapsula's Blog*, 9 December 2013. Available at <http://www.incapsula.com/blog/bot-traffic-report-2013.html>, accessed 23 March 2015.

Defining Autonomy in Systems: Challenges and Solutions

Andrew Williams

Abstract

The subject of ‘autonomy’ or ‘autonomous systems’ is of growing importance in both military and civilian domains. Technology is evolving rapidly and recent military operations increasingly rely on robotic and unmanned systems, which incorporate increasingly ‘autonomous’ functions. Multiple definitions and understandings currently exist about autonomous systems, and the related concepts of autonomy, automation, robotics, and unmanned systems. This terminological confusion makes basic coherence of defence programmes challenging, and at worst, affects strategic level defence policy when certain concepts are misrepresented in public debate. This chapter analyses key terms in the field of autonomous systems, illustrates problems with each, and provides solutions for overcoming conceptual difficulties. It recommends replacing the widely used but ambiguous term ‘autonomous system’ with the more specific ‘system with autonomous functions’.

Introduction

Rigorously specified definitions are essential for systematic knowledge. They facilitate common understanding and meaningful and informed discussion of a subject or phenomenon of interest. This chapter considers definitions in the field of ‘autonomous systems’ and asks several, rather simple questions that have significant implications for how policies and laws are interpreted and understood. First, what are the differences between automatic, autonomic, autonomous, and other commonly encountered terms in this field? Second, how can the characteristics of autonomy be categorised and measured? Are there ‘levels’ of autonomy? Finally, how should ‘autonomous system’ be defined?

This chapter addresses the fact that autonomy is both a difficult concept to understand and is used inconsistently and inappropriately in current policy

discussions. The development and use of military weapon systems involving autonomous functions is undoubtedly the most controversial issue in this debate, and an area in which policy makers should place the highest levels of scrutiny. Current narratives tend to conflate the use of such systems in recent conflicts, with the fact that they may employ autonomous functions. Furthermore, a plethora of terms are used inconsistently and interchangeably – unmanned platform, autonomous systems, robot, drone – and sometimes in combination, such as unmanned autonomous robot.⁽¹⁾ Often, additional descriptors are added to focus the attention of the public and policy makers on a particular issue, for example the recent debates about ‘killer robots’. That the possibility of autonomous weapons warrants extremely close scrutiny from ethical, legal, and technical perspectives is unquestionable. Terminological confusion, however, is obscuring rational debate about the nature of systems with autonomous functions.

This chapter is based on a study conducted from 2013–14 as part of the Multinational Capability Development Campaign (MCDC) – a collaborative program of work between 19 nations, NATO Allied Command Transformation, and the European Union.⁽²⁾ An analytic inductive qualitative research approach was used, which looked for common themes and issues in the literature. A typological analysis⁽³⁾ looking at the fundamental components of autonomy was created. The initial findings were validated at several subject-matter expert workshops and refined based on this input. Data was drawn from primarily publicly available documents, but also from discussion with subject matter experts.

Definitional Analysis

The first question in this chapter considers the differences between commonly encountered terminologies in the field of autonomous systems. This section covers several key terms: unmanned, remote-controlled, robot, system, automatic, autonomic, and autonomy. These key terms are integral to

1. *Register* 2014.

2. Full study findings are described in *Williams* 2014.

3. *Bailey* 1994; *Given* 2008.

understanding key policy and legal issues concerning autonomous systems, in that they describe the core concepts in the subject area.

Review of Key Definitions

Unmanned

‘Unmanned’ is usually used in combination with a system, in particular an unmanned aircraft system (UAS) or unmanned aerial vehicle (UAV), but also unmanned ground or sea platforms. A working group formed under the United States’ National Institute of Standards and Technology (NIST) defined an **unmanned system [platform]**⁽⁴⁾ as:

An electro-mechanical system [platform], with no human operator aboard, that is able to exert its power to perform designed missions. May be mobile or stationary. Includes categories of unmanned ground vehicles (UGVs), UAVs, unmanned underwater vehicles (UUVs), unmanned surface vehicles (USVs), unattended munitions (UMs), and unattended ground sensors (UGSs). Missiles, rockets, and their sub-munitions, and artillery are not considered unmanned systems.

While the general meaning of unmanned is obvious, it is important to distinguish between unmanned *systems* and unmanned *platforms*; many proposed definitions do not do so. Systems are considered as the totality of a capability (hardware, communications links, crew, platform and sensors etc.), while platform refers to the part of the capability conducting operational tasks. A Predator platform may be unmanned, but it is supported by an extensive ground crew and operator, therefore the system is manned. The US Department of Defense distinguishes between system and platform:

- **unmanned aircraft:** an aircraft or balloon that does not carry a human operator and is capable of flight under remote control or autonomous programming;⁽⁵⁾ and

4. NIST 2012. The platform has been added in brackets to avoid confusion.

5. DOD 2010.

- **unmanned aircraft system:** that system whose components include the necessary equipment, network, and personnel to control an unmanned aircraft.

It is important to note that the ‘unmanned’ nature of a platform does not logically lead to that platform operating with some degree of autonomy. Most unmanned platforms rely on some degree of human control.

Remote controlled

Remote control refers to operation of a system or activity by a person at a different place, usually by means of radio or ultrasonic signals or by electrical signals transmitted by wire. It also refers to the device used to control a machine from a distance. The NIST defines ‘**remote controlled**’ as:

A mode of operation of an unmanned system [platform] wherein the human operator, with benefit of video or other sensory feedback, directly controls the actuators of the unmanned system [platform] on a continuous basis, from off the vehicle and via a tethered or radio linked control device using visual line-of sight cues. In this mode, the UxS takes no initiative and relies on continuous or nearly continuous input from the user.⁽⁶⁾

Remote control is associated with autonomy for the reason that unmanned platforms with low levels of autonomy require remote control. It is usually assumed that a remote-controlled platform is also an unmanned one, however, there are cases where some functions of manned platforms are also remote controlled (in the case of many spacecraft platforms). Consequently, the above definition is restrictive, rather than general. More often, remote control is expressed in terms of remote ‘piloting’ such as the formal NATO definition for **remotely piloted aircraft**.⁽⁷⁾ ‘An unmanned aircraft that is controlled from a remote pilot station by a pilot who has been trained and certified to the same standards as a pilot of a manned aircraft.’

6. NIST 2012.

7. NATO 2014.

Robot

Various international standardisation bodies and professional groups are working to create a standardised terminology set for robots, as current definitions tend to mix core characteristics with intended uses (e.g., an autonomous machine that performs tasks in place of humans). Four recent definitions of **robots** include:

1. an actuated mechanism that is programmable in two or more axes and has a degree of autonomy, moving within its environment, to perform intended tasks;⁽⁸⁾
2. an automatically controlled, reprogrammable multi-purpose manipulator that is programmable in three or more axes;⁽⁹⁾
3. a powered physical system designed to be able to control its sensing and action in order to accomplish assigned tasks in the physical environment (including its associated human-robot interface, HRI);⁽¹⁰⁾ and
4. a machine which, through remote control or based on pre-programmed patterns, can carry out tasks of certain complexity with various degrees of autonomy from human supervision. If these tasks involve the use of armed force, they can be described as ‘robotic weapons’ or ‘unmanned weapon systems’.⁽¹¹⁾

As the above definitions are indistinguishable from those of ‘autonomous system’, for the remainder of the chapter and throughout this volume, the terms are used interchangeably.

Automatic

Automatic can be defined in both a conceptual dictionary sense, and in the context of specific systems. The Oxford American English dictionary lists five definitions of **automatic**:

1. operating by itself without human control;
2. (of a gun): able to fire continuously until the bullets run out;
3. done without conscious thought;

8. ISO 2012.

9. This standard, ISO 8373, dated 1996, was developed with industrial service robots in mind.

10. NIST 2012.

11. Melzer 2013.

4. occurring spontaneously: automatic enthusiasm; and
5. (of a punishment): applied without question because of a fixed rule.

There are also several definitions of automatic (automated) in the context of a system:

- **automatic**: a process that may be executed independently from start to finish without any human intervention;⁽¹²⁾
- **automatic system**: a system that has fixed choice points, programmed with a number of fixed alternative actions that are selected by the system in response to inputs from particular sensors;⁽¹³⁾ and
- **automated system** (in the unmanned aircraft context): a system that, in response to inputs from one or more sensors, is programmed to logically follow a pre-defined set of rules in order to provide an outcome; its output is predictable if the set of rules under which it operates is known.⁽¹⁴⁾

Automatic is often used interchangeably with autonomous or autonomic (see below). In many cases this is incorrect usage; however, when considering the functioning of a system on a macro scale, automatic can be considered to belong to the same class of concepts as autonomy. This is elaborated further in the next section.

Autonomic

Autonomic is often used interchangeably with automatic, which is misleading due to fundamental meanings, and autonomous, which is misleading due to common use. A dictionary definition of **autonomic** is: (1) occurring involuntarily or spontaneously; (2) of or relating to the autonomic nervous system; or (3) occurring as a result of internal stimuli.

There are several definitions of autonomic in the context of systems, for example: (1) in electronics, the study of self-regulating systems for process control and (2) of, or pertaining to, the capacity of a system to control its own internal state and operational condition.⁽¹⁵⁾

12. Dunlop 2009.

13. Ibid.

14. MOD 2012.

15. Truszkowski et al. 2009.

Autonomy/autonomous

Like automatic and autonomic, autonomous (adjective) or autonomy (noun) can be defined in both a general sense and relative to a particular context (i.e., system). In a general sense, **autonomous** is defined as:

1. having the quality of being self-governing (of a community, country, etc.); possessing a large degree of self-government;
2. of, or relating to, an autonomous community;
3. independent of others;
4. philosophy: acting, or able to act, in accordance with rules and principles of one's own choosing; compare with heteronomous; see also categorical imperative (in the moral philosophy of Kant, of an individual's will) directed to duty rather than to some other end;
5. biology: existing as an organism independent of other organisms or parts.

In the context of systems, two common definitions of **autonomy** are:

1. A system's capacity to act according to its own goals, precepts, internal states, and knowledge, without outside intervention.⁽¹⁶⁾
2. An unmanned system's [platform's] own ability of integrated sensing, perceiving, analysing, communicating, planning, decision making, and acting/executing to achieve its goals as assigned by its human operators through HRI or by another system with which that the unmanned system communicates. Unmanned system autonomy is characterised into levels from the perspective of Human Independence, the inverse of HRI. Autonomy is further characterised in terms of contextual autonomous capability.⁽¹⁷⁾

An **autonomous system** is capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system can take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present.

16. Truszkowski et al. 2009.

17. NIST 2012.

Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be.⁽¹⁸⁾

System

System is a widely used term that has many potential meanings. In the defence context, it is often used to describe physical military hardware in general (e.g., missile systems), specific platforms (e.g., unmanned aerial vehicle, tank, ship), or component parts (e.g., electronic, mechanical, or computer systems). The general dictionary definitions of **system** are:⁽¹⁹⁾

1. a combination of things or parts forming a complex or unitary whole;
2. any assemblage or set of correlated members; *a system of currency*;
3. an ordered and comprehensive assemblage of facts, principles, and doctrines in a particular field of knowledge or thought: *a system of philosophy*.
4. a coordinated body of methods or a scheme or plan of procedure; organisational scheme: *a system of government*;
5. any formulated, regular, or special method or plan of procedure: *a system of marking, numbering, or measuring*; *a winning system at bridge*.

Given this term's wide use in common business, engineering, and technical parlance, there is a wide variety of context-specific definitions:

1. 'an aggregation of end products and enabling products to achieve a given purpose'.⁽²⁰⁾
2. an integrated composite of people, products, and processes that provides a capability to satisfy a stated need or objective.⁽²¹⁾
3. a functionally, physically, and/or behaviourally related group of regularly interacting or interdependent elements that forms a unified whole.⁽²²⁾
4. an organised, purposeful structure that consists of inter-related and interdependent elements (components, entities, factors, members, parts, etc.). These elements continually influence one another (directly or

18. MOD 2012.

19. <http://dictionary.reference.com/browse/system>, accessed 25 March 2015.

20. ANSI 1999.

21. DOD 2001.

22. DOD 2010.

indirectly) to maintain their activity and the existence of the system in order to achieve the goal of the system.⁽²³⁾

5. a set or arrangement of related elements and processes, the behaviour of which satisfies customer/operational needs and provides for life-cycle sustainment of the products.⁽²⁴⁾
6. a combination of elements that function together to produce the capability to meet a need. These elements include all hardware, software, equipment, facilities, personnel, processes, and procedures needed for this purpose.⁽²⁵⁾
7. a construct or collection of different elements that together produce results not obtainable by the elements alone. The elements, or parts, can include people, hardware, software, facilities, policies, and documents (i.e., all things required to produce systems-level results). The results include system-level qualities, properties, characteristics, functions, behaviour, and performance. The value added by the system as a whole, beyond that contributed independently by the parts, is primarily created by how the parts are interconnected.⁽²⁶⁾

These definitions indicate that a system is always a composite entity, but they do not provide guidance on where to draw the boundary. System is a scalable concept. The following list demonstrates the nested nature of different levels of systems. Each level is broader than the previous level, and includes more components:

- electronic transistor-microchip system;
- flight control system;
- airplane system;
- airplane-crew system;
- airplane-ground control system;
- integrated air defence system; and
- national defence system.

23. <http://www.businessdictionary.com/definition/system.html#ixzz2dBfpOrSj>, accessed 25 March 2015.

24. IEEE 1998.

25. NASA 2007.

26. International Council on Systems Engineering. Available at <http://www.incose.org/practice/fellows-consensus.aspx>, accessed 25 March 2015.

Other definitions of importance

Other definitions that are likely important in the discussion of autonomous systems are beyond the scope of this research. In the legal domain, for example, critical terms are likely to include: autonomous weapon, autonomous weapon system, pilot, operator, programmer, and programming station. The weaponisation of autonomous systems is a separate subject and of considerable legal importance.⁽²⁷⁾ Furthermore, understanding legal responsibility in harm situations may depend on who is in control, which necessitates distinguishing between a 'pilot' and 'programmer'.

Another widely used term is 'drone'. However, none of the official sources consulted in the review conducted for this work offered a definition of the term.⁽²⁸⁾ Instead, 'drone' has now become synonymous with weaponised unmanned air systems or remotely piloted aircraft and their use in recent conflicts.⁽²⁹⁾ While the term certainly garners attention in the media, and is commonly used by advocacy groups to protest the ethical issues of arming unmanned systems and the use of such systems – remotely controlled or autonomous – in certain operational situations, it is too broad to be meaningful.

Analysis

Definitional analyses are conducted to develop a clear understanding of a subject area; reduce ambiguity and the potential for confusion; and to select or develop a set of useful, minimal, and mutually exclusive definitions for a policy area. The following discussion considers the key terms and their relations to one another.

Understanding of 'system'

As previously noted, it is potentially confusing to use 'system' to refer to a specific platform (e.g., a drone), because the platform would likely be tightly

27. For a recent analysis of this topic, see Scharre and Horowitz 2015.

28. In older versions of the NATO glossary of terms and definitions (AAP-6) document, drone was defined as 'an unmanned vehicle which conducts its mission without guidance from an external source'. This definition was deleted in the most recent version of AAP-6 (NATO 2014).

29. For further discussion on this topic, see <http://www.dronedefinition.com/3-new-names-for-drone-definition/>, <http://dronewars.net/aboutdrone/>, and <http://blogs.scientificamerican.com/guest-blog/2012/04/12/what-is-a-drone-anyway/>, accessed 25 March 2015.

coupled to a control station with personnel. As the definitions of system above indicate, the conventional understanding in defence is to consider ‘system’ to be the total capability, rather than the specific part that performs operational duties. However, it is important to be consistent about the scale within any particular project: if the term ‘autonomous system’ is used, the user must clearly define what they mean by ‘system’.

Use of ‘unmanned’ in an autonomous systems context

It is misleading to focus on the ‘unmanned’ nature of a platform, especially in the context of autonomy. Many platforms are unmanned but completely controlled by a crew, while others are unmanned but only partially controlled by a crew. Some platforms are manned, but delegate certain functions to a crew in another location (e.g., spacecraft).

Platforms – the part of a system that executes operational tasks – perform a variety of functions, and may be manned or unmanned. These functions belong to the larger set performed by the relevant system to which the platform belongs. Overall, some of these functions may be under human control; others may not.

From the standpoint of autonomy, it is more fundamental to ascertain what *functions* a system performs (and the extent to which these functions are autonomous or under human control) than to ask whether a human is physically located with any particular part of the system. Autonomy confers a range of benefits *in addition* to ‘unmannedness’: faster execution of tasks, higher interoperability, lower error rates, and the increasing use of unmanned platforms, which decreases risk to life and improves operational access. Thus in addition to studying the implications of increasing the autonomy of unmanned platforms, the autonomisation of various functions of *manned* systems should also be explored.

Understanding of ‘autonomous system’

The above discussion addresses the potential confusion surrounding the term ‘autonomous system’. As noted, the preferred definition of ‘systems’ in the defence field refers to a composite of platform, controllers, enabling infrastructure, networks, etc. It is impossible to have a completely autonomous system, because defence systems are generally embedded with some level of human

control. In strict terms, a ‘fully autonomous system’ would be in complete control of itself, with no human intervention at all. Thus ‘levels of autonomy’ refer to a transfer of control from one part of the system (e.g., ground control with humans) to another (e.g., the platform executing operational tasks).

An example of an ‘autonomous platform’ would be an autonomous tactical drone that is under overall mission control from a controller in the system. However, this is unnecessarily complicated and creates a conceptual paradox regarding where autonomy starts and ends in system terms. To avoid this problem, system boundaries should be explicitly defined in the context of the ‘role of autonomy’. For example:

- *The designers of a reconnaissance system have autonomised the flight, target selection, and data-processing functions of a reconnaissance plane (it is irrelevant whether it is manned or unmanned). This statement makes clear that only parts of the system are autonomous.*
- *An improvised explosive device (IED) disarming system is autonomous, within certain high-level constraints set by the control station staff who monitor the disarming system as part of an integrated IED defence system. This statement considers the IED disarmer as a system, but is clear that it is part of a larger system that includes human control.*
- *A smart micro-unmanned aerial target selection system is completely autonomous after being given initial instructions. It optimises its search parameters based on artificial intelligence algorithms and the current terrain and meteorological conditions, and maximises data output when close to ground receivers. It is programmed to self-destruct upon loss of power and is intended to be non-recoverable. This statement defines the system as the micro-unmanned aerial sensor, and it is clear that the system is completely autonomous.*

Automatic vs. autonomous

There are several core features of the definitions of automatic. First, ‘automatic’ implies an unconscious process that lacks actual decision making. Second, apparent decision choices are ‘simulated’ by the system, which follows a pre-defined and fixed set of rules to translate a restricted set of inputs into a deterministic and limited range of outputs. Third, once started, it proceeds without human intervention and cannot deviate from its restricted programming.

In contrast, *autonomy* implies self-government, self-management, and decision making. Decision choices are carried out according to pre-defined conditions or system constraints. Outputs are non-deterministic, or probabilistic, and dependent on detected changes in the environment, which are random.

Automatic and autonomous can be confused because autonomous behaviour can be characterised at a macro level as automatic because no human intervention is required. Given a certain set of inputs, autonomous systems can operate ‘automatically’ without intervention. Yet this usage misses the key point: unlike automatic systems, autonomous ones will not always yield the same outputs under the same inputs and conditions.⁽³⁰⁾

In summary, from the perspective of determining the appropriate amount of human control, the difference between automatic and autonomous is almost irrelevant; the important issue is the level of human control over a system. From the perspective of understanding what autonomy is, however, it is necessary to make this distinction.

Autonomic vs. autonomous

The origin of the word autonomic is a combination of ‘autonomous’ and the suffix ‘-ic’. Adding ‘ic’ to a word means ‘of or pertaining to’ a particular characteristic. Formally, autonomic should be used always in reference to a particular concept or object – i.e., autonomic nervous system. However, its use applies to autonomy in the context of *internal* self-regulation, as opposed to autonomy in acting in the *external* environment.

Levels of Autonomy

The second question addressed by this chapter asks how the characteristics of autonomy can be categorised and measured. This section reviews previous attempts to define ‘levels’ or ‘scales’ of autonomy. A key reason for creating

30. A critique of this position is that, given that the output probability space (rather than a discrete parameter set) is known for autonomous processes, these processes could be considered very complicated *automatic* processes. If the decision-making algorithms were only linear programming optimisation and stochastic Monte Carlo types, then it may be possible to consider the processes as automatic, even if only the output probability distributions were known, rather than discrete parameters. It is unlikely, however, that useful systems could be designed using this class of decision algorithm. To be truly self-governing, autonomous systems must employ evolutionary and adaptive decision-making algorithms, which allow them to learn and adapt their behaviour, meaning that while the overall behaviour could be predictable, individual actions are not.

levels of autonomy was to help defence leadership and system designers track the progress of defence technology programmes and determine whether goals to increase the autonomy of certain systems had been met.⁽³¹⁾ As will be seen, however, this approach became a key way to try and understand both what autonomy is and its implications for defence systems. There are three distinct ways in the literature to construct levels of autonomy: categorical linear scales, multidimensional scales, and contextual scales.

Categorical Linear Scales

Early attempts to create levels of autonomy used simple linear scales that were categorical as opposed to cardinal (i.e., there is no meaning to the ‘distance’ between levels, such that the difference between 1 and 2 cannot be compared with the difference between 7 and 8, nor is level 6 twice as autonomous as level 3).

Sheridan, who was then working for NASA, created an early scheme: a 10-level scale of degrees of automation, which incorporate autonomy.⁽³²⁾ Many of Sheridan’s examples focus on telerobotics, in which the human is physically separated from the system but is still issuing commands. Levels 2–4 are centred on who makes the decisions, the human or the computer. Levels 5–9 are centred on how to execute that decision. In particular, Sheridan provides insight into the effects that an operator’s trust (or lack thereof) has on an autonomous system. Trust in autonomy comprises both the perceived technical and operational reliability and the confidence that a human places in autonomy.

31. Clough 2000.

32. Research into levels of autonomy often references Sheridan (1992).

Table 2.1. Sheradin's Scale of Automation

-
- (1) The computer offers no assistance; the human must do it all.
 - (2) The computer offers a complete set of action alternatives, and...
 - (3) narrows the selection down to a few, or...
 - (4) suggests one, and...
 - (5) executes that suggestion if the human approves, or...
 - (6) allows the human a restricted time to veto before automatic execution, or...
 - (7) executes automatically, then necessarily informs the human, or...
 - (8) informs him after execution only if he asks, or...
 - (9) informs him after execution if it, the computer decides to.
 - (10) The computer decides everything and acts autonomously, ignoring the human.
-

A similar approach is used by the US Navy Office of Naval Research and the UK's Systems Engineering for Autonomous Systems Defence Technology Centre.⁽³³⁾

Table 2.2. US Navy Office of Naval Research Levels of Autonomy

Level	Name	Description
1	Human operated	All the activity in the system is a direct result of human-initiated environment, although it may have information-only responses to sensed data.
2	Human assisted	The system can perform activity in parallel with human input, acting to augment the human's ability to perform the desired activity, but has no ability to act without accompanying human input. An example is automobile automatic transmission and anti-skid brakes.
3	Human delegated	The system can perform limited control activity on a delegated basis. The level encompasses automatic flight controls, engine controls, and other low-level automation that must be activated or deactivated by a human input and act in mutual exclusion with human operation.

33. Williams 2008.

Level	Name	Description
4	Human supervised	The system can perform a wide variety of activities given top-level permissions or direction by a human. The system provides sufficient insight into its internal operations and behaviours that it can be easily understood by its human supervisor and appropriately redirected. The system does not have the capability to self-initiate behaviours that are not within the scope of its current directed tasks.
5	Mixed initiative	Both the human and the system can initiate behaviours based on sensed data. The system can coordinate its behaviour both explicitly and implicitly. The human can understand the behaviours of the system in the same way that he or she understands his or her own behaviours. A variety of means are provided to regulate the authority of the system with respect to human operators.
6	Fully autonomous	The system requires no human intervention to perform any of its designed activities across all planned ranges of environmental conditions.

Table 2.3 shows another level of autonomy scale, which was developed by Clough⁽³⁴⁾ for UAVs. As the levels get higher, the scope and complexity of the UAV's independent behaviour increase. It is clear that there are different challenges at different levels. It is assumed that a UAV with lower levels of autonomy would be easier to prove safe. It can also be noted that the first few levels of the Clough scale, up to and including level 4, denote the system's increasing ability to autonomously resolve situations. There is also the potential for reduced operator workload, and the detection and querying of incorrect human commands.

One question that arises is whether a vehicle that is fully dependent on a remote controller can ever truly be 'fail safe' if communication is lost. A system that has no further autonomous capabilities cannot perform any contingency procedures (such as returning to base). It can be assumed that the higher levels of autonomy may introduce new hazards related to their new capabilities. For example, level 7 has 'predictive battlespace data in limited range', which could allow it to attack a friendly position on the basis of inaccurate predictions of friendly and enemy involvement.

34. Clough 2002.

Table 2.3. Clough's Levels of Autonomy

Level	Description
0	Remotely piloted vehicle
1	Execute pre-planned mission
2	Changeable mission
3	Robust response to real-time faults/events
4	Fault/event adaptive vehicle
5	Real-time multi-vehicle coordination
6	Real-time multi-vehicle cooperation
7	Battlespace knowledge
8	Battlespace cognizance
9	Battlespace swarm cognizance
10	Fully autonomous

In another approach, a study group for the NATO Industrial Advisory Group (NIAG) conducted a pre-feasibility study on autonomous operations⁽³⁵⁾ to obtain different industries' views on what degree of autonomy is achievable for safe autonomous operations in the future. NIAG identified the degree of autonomy that was feasible and the technologies required, and assessed the maturity of enabling technologies. The study also considered several systems analysis methodologies that could be applied in the human behavioural process in unmanned systems. The working group decided to use the observe, orient, decide, act (OODA) loop (Figure 2.1) to map the behaviour processes against the capabilities. It identified four levels of autonomy:

- Level 1: remotely controlled system: system reactions and behaviour depend on operator input (non-autonomous);
- Level 2: automated system: reactions and behaviour depend on fixed built-in functionality (pre-programmed);
- Level 3: autonomous non-learning system: behaviour depends upon fixed built-in functionality or a fixed set of rules that dictates system behaviour (goal-directed reaction and behaviour); and
- Level 4: autonomous learning system with the ability to modify rule-defining behaviours: behaviour depends upon a set of rules that can be

35. NIAG 2004.

modified to continuously improve goal-directed reactions and behaviours within an overarching set of inviolate rules/behaviours.

Using these definitions, the NIAG working group determined that nearly all of today's UAV systems were either level 1 or 2. According to their definition, to be autonomous, a system must have the ability to operate without continual human intervention. Consequently, according to NIAG, autonomy does not eliminate the human in the loop but instead changes his or her level of interaction with the system.

There are several advantages of using simple categorical linear scales to assess the level of autonomy. They give policy makers a reasonable conceptual

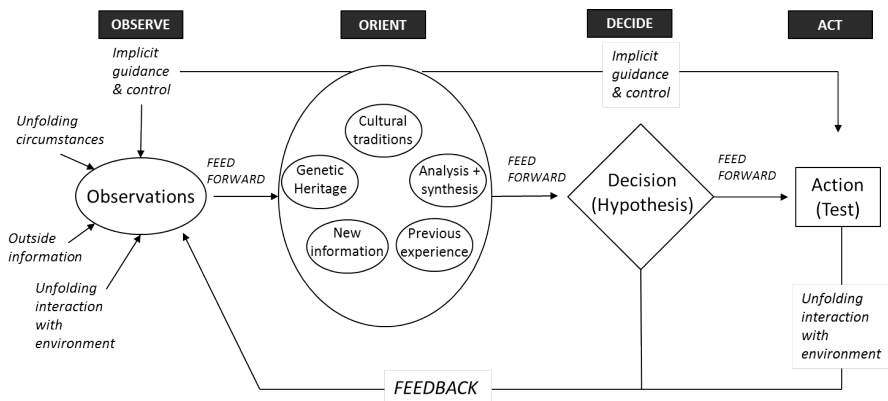


Figure 2.1. John Boyd's OODA Loop

understanding of autonomy and human-machine interface issues, and are simple to incorporate into policy/legal/regulative guidance. The disadvantages, however, are that they are non-empirical and ambiguous, which makes them challenging for systems engineering or technical uses, and they are generic and do not reference specific system functions.

Multi-dimensional Scales

Multi-dimensional scales consist of increasing graduations or levels with descriptive indicators for a set of dimensions that defines a particular level of autonomy. In 2000, Parasuraman et al.⁽³⁶⁾ proposed a revised model to describe levels of automation with four dimensions that categorised the tasks of humans and systems: information acquisition, information analysis, decision and actions selection, and action implementation. While their focus was on the *computer's* decision-making process, Proud et al.⁽³⁷⁾ developed a four-tier system based on the *human* decision-making process. It mapped eight levels of autonomy over four dimensions represented by Boyd's (OODA) loop (Figure 2.1). This approach covers the multi-UAV operations that are absolutely dependent on the implementation of appropriate autonomy levels.

Developed for the NASA Johnson Space Center, Proud et al.'s work aims to find out how autonomous a system should be. After establishing descriptors of levels of autonomy and determining how to categorise the function types, they created an eight-level scale of autonomy for each OODA category to determine the level of autonomy of a particular function.

Proud et al.'s 'Level of Autonomy Assessment Scale' was derived from research into external autonomy applications (Sheridan, Parasuraman, etc.) and other studies. The scale's intention is to help system designers more easily and accurately identify the appropriate level of autonomy at which to design each function of their system. The advantage of employing the OODA category aspect is that it allows more nuanced description than previous scales, and the function type can be weighted differently across a particular level.

Similar to Sheridan, Proud et al.'s level of autonomy scale is bounded by levels 1 and 8, which correspond to complete responsibility for the human and computer, respectively. The team tailored each level of autonomy scale to fit the tasks by function type (OODA). The levels of autonomy were broken down into three sections. In levels 1–2, the human is primary; in levels 3–5, the computer operates with human interaction; and in levels 6–8, the computer operates independently of the human, who has decreasing access to information and over-ride capabilities.

36. Parasuraman, Sheridan, and Wickens 2000.

37. Proud, Hart, and Mrozinski 2003.

Table 2.4. Proud et al.'s Level of Autonomy Assessment Scale

Level	Observe	Orient	Decide	Act
8	The computer gathers, filters, and prioritises data without displaying any information to the human.	The computer predicts, interprets, and integrates data into a result that is not displayed to the human.	The computer performs ranking tasks. The computer performs final ranking, but does not display results to the human.	Computer executes automatically and does not allow any human interaction.
7	The computer gathers, filters, and prioritises data without displaying any information to the human, though a 'program functioning' flag is displayed.	The computer analyses, predicts, interprets, and integrates data into a result that is only displayed to the human if it fits the programmed context (context-dependent summaries).	The computer performs ranking tasks. The computer performs final ranking and displays a reduced set of ranked options without displaying 'why' decisions were made to the human.	Computer executes automatically and only informs the human if required by context. It allows for over-ride ability after execution. Human is shadow for contingencies.
6	The computer gathers, filters, and prioritises information displayed to the human.	The computer overlays predictions with analysis and interprets the data. The human is shown all results.	The computer performs ranking tasks and displays a reduced set of ranked options while displaying 'why' decisions were made to the human.	Computer executes automatically, informs the human, and allows for over-ride ability after execution. Human is shadow for contingencies.
5	The computer is responsible for gathering the information for the human, but it only displays non-prioritised, filtered information.	The computer overlays predictions with analysis and interprets the data. The human shadows the interpretation for contingencies.	The computer performs ranking tasks. All results, including 'why' decisions were made, are displayed to the human.	Computer allows the human a context-dependent restricted time to veto before execution. Human shadows for contingencies.
4	The computer is responsible for gathering the information for the human and for displaying all information, but it highlights the non-prioritised, relevant information for the user.	The computer analyses the data and makes predictions, though the human is responsible for interpreting it.	Both the human and computer perform ranking tasks, and the results from the computer are considered prime.	Computer allows the human a pre-programmed restricted time to veto before execution. Human shadows for contingencies.

Level	Observe	Orient	Decide	Act
3	The computer is responsible for gathering and displaying unfiltered, unprioritised information for the human. The human is still the prime monitor of all information.	Computer is the prime source of analysis and predictions, with a human shadow for contingencies. The human is responsible for interpretation of the data.	Both the human and computer perform ranking tasks, and the results from the human are considered prime.	Computer executes decision after human approval. Human shadows for contingencies.
2	Human is the prime source for gathering and monitoring all data, with computer shadow for emergencies.	Human is the prime source of analysis and predictions, with computer shadow for contingencies. The human is responsible for interpretation of the data.	The human performs all ranking tasks, but the computer can be used as a tool for assistance.	Human is the prime source of execution, with computer shadow for contingencies.
1	Human is the only source for gathering and monitoring (defined as filtering, prioritising, and understanding) all data.	Human is responsible for analysing all data, making predictions, and interpreting the data.	The computer does not assist in or perform ranking tasks. Human must do it all	Human alone can execute decision.

In a similar approach, Clough developed an Autonomy Control Level (ACL) framework. The ACL specifies 11 levels of autonomy (which are mapped across OODA loop dimensions) and introduces the notion that higher levels are defined by the ability to interact with multiple platforms.

Table 2.5. Clough's Autonomy Control Level Framework

Level	Level Descriptor	Observe Perception / Situational Awareness	Orient Analysis / Coordination	Decide Decision Making	Act Capability
10	Fully autonomous	Cognizant of all within battlespace	Coordinates as necessary	Capable of total independence	Requires little guidance to do job

Level	Level Descriptor	Observe Perception / Situational Awareness	Orient Analysis / Coordination	Decide Decision Making	Act Capability
9	Battlespace swarm cognizance	Battlespace inference – intent of self and others (allies and foes) Complex/intense environment – on-board tracking	Strategic group goals assigned Enemy strategy inferred	Distributed tactical group planning Individual determination of tactical goal Individual task planning/execution Choose tactical targets	Group accomplishment of strategic goal with no supervisory assistance
8	Battlespace cognizance	Proximity inference – intent of self and others (allies and foes) Reduced dependence on off-board data	Strategic group goals assigned Enemy tactics inferred ATR	Coordinated tactical group planning Individual task planning/execution Choose targets of opportunity	Group accomplishment of strategic goal with minimal supervisory assistance (example: go SCUD hunting)
7	Battlespace knowledge	Short track awareness – history and predictive battlespace data in limited range, timeframe, and numbers Limited inference supplemented by off-board data	Tactical group goals assigned Enemy trajectory estimated	Individual task planning/execution to meet goals	Group accomplishment of tactical goal with minimal supervisory assistance
6	Real-time multi-vehicle cooperation	Ranged awareness – on-board sensing for long range, supplemented by off-board data	Tactical group goals assigned Enemy location sensed/estimated	Coordinate trajectory planning and execution to meet goals –group optimisation	Group accomplishment of tactical goal with minimal supervisory assistance Possible close airspace separation (1-100 yds)

Level	Level Descriptor	Observe Perception / Situational Awareness	Orient Analysis / Coordination	Decide Decision Making	Act Capability
5	Real-time multi-vehicle coordination	Sensed awareness – local sensors to detect others Fused with off-board data	Tactical group plan assigned Real-time health diagnosis; ability to compensate for most failures and flight conditions; ability to predict onset of failures Group diagnosis and resource management	On-board trajectory re: planning – optimises for current and predictive conditions Collision avoidance	Group accomplishment of tactical plan as externally assigned Air collision avoidance Possible close airspace separation for AAR, formation in non-threat conditions
4	Fault/event adaptive vehicle	Deliberate awareness – allies communicate data	Tactical plan assigned Assigned rules of engagement Real-time health diagnosis; ability to compensate for most failures and flight conditions – an inner-loop changes reflected in outer-loop performance	On-board trajectory re: planning – event driven Self-resource management DE confliction	Self-accomplishment of tactical plan as externally assigned
3	Robust response to real-time faults/events	Health/status history and models	Tactical plan assigned Real-time health diagnosis (what is the extent of problems?) Ability to compensate for most control failures and flight conditions (i.e., adaptive inner-loop control)	Evaluate status vs. requirement mission capabilities Abort/RTB if insufficient	Self-accomplishment of tactical plan as externally assigned

Level	Level Descriptor	Observe Perception / Situational Awareness	Orient Analysis / Coordination	Decide Decision Making	Act Capability
2	Changeable mission	Health/status sensors	Real-time health diagnosis (do I have problems?) Off-board replan (as required)	Execute pre-programmed or uploaded plans in response to mission and health conditions	Self-accomplishment of tactical plan as externally assigned
1	Execute pre-planned mission	Pre-loaded mission data Flight control and navigation sensing	Pre-/post-flight BIT Report status	Pre-programmed mission and abort plans	Wide airspace separation requirements (miles)
0	Remotely piloted vehicle	Flight control (altitude, rates) sensing Nose camera	Telemetered data Remote pilot commands	N/A	Control by remote pilot

Multi-dimensional functional scales are useful for systems engineering including cost-benefit or system trade-off calculations about levels of autonomy, but they have several disadvantages. First, the interpretation of any given level of autonomy depends strongly on the functional division and classification, meaning that the level may be affected by how the functions are described. Second, it is unclear how to deal with situations in which a level in one particular dimension may be different from that in another dimension. There may be multiple possible combinations, defined as the number of possible ways to choose d from $[l \times d]$, where d = the number of dimensions and l = the number of levels.

Finally, these scales do not help define autonomy, as there is no ‘intrinsic’ scale of autonomy that is meaningful only in reference to a given system function. Furthermore, like the linear scales, the multi-dimensional scales tend to mix graduations of automation and autonomy, thus confusing the distinction between the two.

Contextual Scales

A third set of scales defines levels of autonomy in context of (or in reference to) a set of conditions external to the system. The NIST set up a group

to address the autonomy issue, which proposed a framework to consider the Autonomy Levels for Unmanned Systems (ALFUS).⁽³⁸⁾ This framework defined the autonomy of an unmanned system as its ability to achieve its mission goals: the more complex the goal, the higher its level of autonomy. Several other studies have been conducted in the United States, Australia, and Europe, which have generated alternative metrics with which to assess the quantitative and qualitative aspects of the autonomy of a system. The ALFUS created a detailed model to characterise the levels of autonomous unmanned systems and their missions along three axes: mission complexity, human interface, and environmental difficulty.

The mission is decomposed into sub-tasks, and a metric score is assigned to each of the three axes. Moving from the lowest to the highest level, a metric score and weight is assigned to each axis for a given sub-task. The metric scores and weights are averaged for each sub-task level and combined with the next-higher level's score.

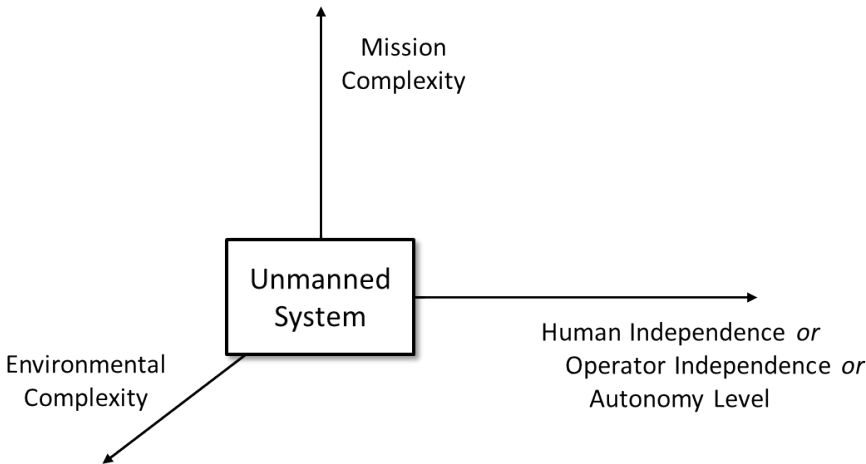


Figure 2.2. Autonomy Level for Unmanned Systems Framework

38. NIST 2012.

A more recent attempt to define the levels of autonomy was undertaken by a US Department of Defense (DoD) Defence Science Board Task Force,⁽³⁹⁾ which suggested a completely different approach. It reviewed many DoD-funded studies on levels of autonomy (including the studies referenced above) and concluded that they were not useful to the autonomy design process because they focused too much attention on the computer and trying to define autonomy – and not enough on the collaboration between the computer, systems, and their operators. The study concluded that at any stage of a mission it was possible for a system to occupy more than one level of autonomy at the same time. The task force therefore recommended that the levels of autonomy approach should be replaced with a three-facet autonomous systems framework that is based on cognitive echelon, mission timelines, and human–machine system trade space, which explicitly:

- focuses design decisions on the explicit allocation of cognitive functions and responsibilities between the human and the computer to achieve specific capabilities;
- recognises that these allocations may vary by mission phase as well as by echelon; and
- makes visible the high-level system trades inherent in the design of autonomous capabilities.

Defining Autonomous System?

Finally, this chapter discusses how to define ‘autonomous system’. One of the major challenges in doing so is the competing descriptions of autonomy from either a general sense, or in reference to a particular context or system. In terms of the general quality or character of autonomy, there is an ongoing debate in various scientific disciplines about how to characterise a machine as autonomous – or how to define what makes sentient biological creatures different from machines. Indeed, some analysts argue that attempting to define ‘autonomous systems’ should be avoided, because by definition, machines cannot be autonomous.⁽⁴⁰⁾

39. DOD 2012.

40. See Stensson and Jansson (2014) for an elaboration of this argument.

Table 2.6 elaborates on the conceptual understanding of autonomy *in general* with reference to an ‘agent’. This table was derived from the various definitions in the literature and reviewing concepts such as machine learning, neural networks, and philosophical and political science conceptions of autonomy and self-governance. The table was then discussed at length in a subject-matter expert workshop for the MCDC project, which did not produce a consensus on how best to describe the characteristics of an autonomous system.

Part of the challenge is that the meaning of autonomy can only be conceptualised in detail in relation to conditions that are external to the agent. Autonomy is not a property of a system, such as colour, mass, or temperature. It is instead a relational property that can only be properly defined in a particular context in reference to certain characteristics. Furthermore, several other problems are revealed when a ‘thing’ is labelled as autonomous or assumed to have human-like characteristics. As described by Scharre in the previous chapter of this volume, autonomy is often referred to in the context of three different issues: the human–machine command-and-control relationship, the complexity of the machine, and the type of decision being automated or ‘autonomised’. This section elaborates on some of these problems, and proposes a definition that resolves them.

Table 2.6. Dimensions of Autonomy

Autonomy dimension	Definition
Goals	An autonomous agent has goals that drive its behaviour.
Sensing	An autonomous agent senses both its internal state and the external world by taking in information (e.g., electromagnetic waves, sound waves).
Interpreting	An autonomous agent interprets information by translating raw inputs into a form usable for decision making.
Rationalising	An autonomous agent rationalises information against its current internal state, external environment, and goals using a defined logic (e.g. optimisation, random search, heuristic search), and generates courses of action to meet goals.
Decision making	An autonomous agent selects courses of action to meet its goals.

Autonomy dimension	Definition
Evaluating	An autonomous agent evaluates the consequences of its actions in reference to goals and external constraints.
Adapting	An autonomous agent adapts its internal state and functions of sensing, interpreting, rationalising, decision making, and evaluating to improve its goal attainment.

Ascribing human characteristics to systems

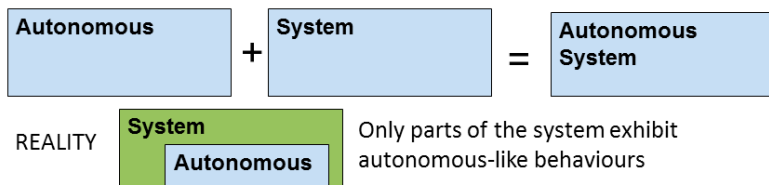
There is a subtle, but vitally important, problem with ascribing human characteristics to a machine. When human concepts such as autonomy, intelligence, or emotion are used as qualifiers for machines – autonomous robot, intelligent platform etc. – several points may be confused.

First, the true meaning of ‘autonomy’ – as described in Table 2.6 – simply may not be understood. Certainly in the defence sector, there is evidence that ‘autonomous system’ is rapidly becoming the term *du jour*. While many users may understand the notion that *machines cannot be autonomous* in the same sense as humans – and can only exhibit ‘autonomous-like’ behaviours relative to a certain level of human control and task-environment complexity – many others do not, and thus either use the term inappropriately or are unnecessarily cautious.

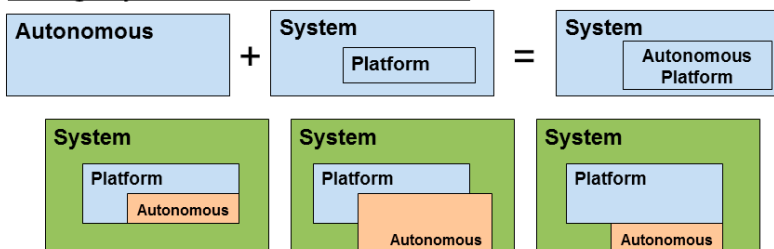
Second, given that a machine or system in totality cannot be autonomous, the term ‘autonomous system’ is used with the unstated assumption that not all parts of the system exhibit autonomous-like behaviour. This creates scope for ambiguity, as any or all of potentially millions of system functions could exhibit autonomous-like behaviours. As described below, modifications such as ‘autonomous platform in a human-machine system’ do not resolve the issues. This confusion is illustrated in Figure 2.3.

While these confusions are partially acknowledged by using terms such as ‘system with autonomous capabilities’, there is still an implication that autonomy is a quality or characteristic *possessed by* the system. The proposed definition below thus reverses this relationship to ‘autonomous functioning within a system’ thus removing the implied grammar of possession.

Ambiguity in “Autonomous System”



Ambiguity in “Autonomous Platform”



REALITY – Autonomous functions are either: exhibited only by part of the platform, or are distributed between platform or system

Figure 2.3. Terminological Ambiguities

Over-applicability of definitions

A major problem of trying to derive a systems definition from the fundamental meaning of autonomy in Table 2.6 is that it can apply equally to sentient creatures (animals or insects) and machines. Thus the suggestion was made by computer scientists in attendance at the MCDC workshops to include the idea that computer programming interacts with the external environment to produce autonomous-like behaviours.

Resolving the definitional problems

The findings of this report emphasise that creating a definition for ‘autonomous system’, ‘autonomous platform’, etc. is inherently misleading without appropriate caution. Ideally, ‘autonomous system’ should be purged from our lexicons; however, this is an impractical suggestion given the prevalent usage of the term. The key point is that using ‘autonomous’ + [system/platform/robot/machine etc.] detracts from the real issue that is most relevant to policy, legal,

and engineering issues: the level of human control necessary and possible over a machine.

Furthermore, such usage singles out ‘autonomy’ as the main descriptor of the machine – over and above all of its other features and capabilities. For example, a particular type of communication may legitimately serve to describe a machine, i.e., ‘radio-enabled’ platforms. When radios were first introduced this may have been logical; however, with the wide use of radio in practically all systems, it is now not appropriate. A similar situation will most likely eventually arise with ‘autonomy-enabled systems’ – eventually their prevalence will be such that this designation seems nonsensical.

Therefore, policy makers and other relevant personnel should not be dissuaded from *generally* using the term ‘autonomous system’, but should be fully aware of the implications, and should caveat documents, presentations, and speeches with the text included in the proposed definition below. Attempting to create definitions for ‘autonomous systems’ should be avoided, because by definition, machines cannot be autonomous.⁽⁴¹⁾

Proposed Definition: ‘Autonomous Functioning in a System’

In light of the recommendation made to avoid using the term ‘autonomous system’, or to at least understand its limitations, a proposal is made to be clearer about the use of autonomy in the context of a system by using instead the term ‘autonomous functioning in a system’. When referring to a particular system or platform, the term ‘system with autonomous functions’ could be used. Table 2.7 provides justification for each part of this definition.

Autonomous Functioning in a System: Autonomous functioning refers to the ability of a system, platform, or software to complete a task without human intervention, using behaviours resulting from the interaction of computer programming with the external environment. Tasks or functions executed by a platform, or distributed between a platform and other parts of the system, may be performed using a variety of behaviours, which may include reasoning and problem

41. See Stensson and Jansson (2014) for an elaboration of this argument.

solving, adaptation to unexpected situations, self-direction, and learning. Which functions are autonomous – and the extent to which human operators can direct, control, or cancel functions – is determined by system design trade-offs, mission complexity, external operating environment conditions, and legal or policy constraints. This can be contrasted with automated functions, which (although they require no human intervention) operate using a fixed set of inputs, rules, and outputs, the behaviour of which is deterministic and largely predictable. Automatic functions do not permit the dynamic adaptation of inputs, rules, or outputs.

Table 2.7. Explanation of Proposed Definition

Component of Definition	Derivation/Understanding
Autonomous Functioning in a System	Conveys the understanding that technology is very far from producing autonomy to the same extent as a sentient animal (i.e., human). It is more realistic to presume that any given defence system will perform a variety of functions, not all of which are autonomous.
...the ability of a system, platform, or software, to complete a task...	Ensures that the difference between system and platform is recognised. Furthermore, in many cases the computer software is the root of an autonomous function, thus this should be recognised in the definition.
...without human intervention...	A key criterion of autonomy.
... using behaviours resulting from the interaction of computer programming with the external environment.	Many definitions of autonomy or autonomous system, including the three listed in the first section of this chapter, could refer to an animal. Thus the definition should differentiate between machines and animals; the key difference is that regardless of how autonomy is conceptualised, it eventually boils down to a computer program interacting in some way with the external environment.

Component of Definition	Derivation/Understanding
Tasks or functions executed either by a platform or distributed between a platform and other parts of the system...	Recognises that autonomy may allow more efficient distribution of functions between various parts of the system.
...which may include...	A rigorous description of <i>behaviour</i> is a very complex task, which involves drawing on a variety of academic disciplines. This phrase was chosen to give flexibility and to convey that autonomous behaviour must include <i>at least one</i> item from the following list.
...reasoning and problem solving, adaptation to unexpected situations, self-direction, and learning	<p>These are drawn from the definitional and dimensional analysis of autonomy.</p> <p>While earlier definitions required the total set of characteristics to be considered autonomous, discussion from workshops revealed that a more flexible definition was preferable.</p> <p>Caution must be used when considering learning. Learning-like behaviour may be observed in system, but this may not actually be learning in the same way that sentient creatures learn. Further work is required to rigorously specify the distinction and overlaps between adaptation and learning.</p>
Which functions are autonomous, and the extent to which human operators can direct, control, or cancel functions, is determined by...	<p>Again, this attributes autonomy to system functions.</p> <p>Makes clear that human operators are able to direct, control, or cancel system functions, depending on certain factors.</p>
...system design trade-offs, mission complexity, external operating environment conditions, and legal or policy constraints.	Caveats the extent of human control and notes that it could vary quite considerably depending on factors <i>external</i> to the system.

Component of Definition	Derivation/Understanding
<p>This can be contrasted against automated functions, which although they require no human intervention, operate using a fixed set of inputs, rules, and outputs, the behaviour of which is deterministic and largely predictable. Automatic functions do not permit dynamic adaptation of inputs, rules, or outputs.</p>	<p>This is an optional caveat to the definition, which contrasts autonomy with automatic. It suggests the key difference between automated and autonomous: while both occur without human intervention, automated functions are to a large extent ‘fixed’, while autonomous functions have some dynamic adaptability. This also suggests that the majority of defence systems that are currently labelled as autonomous would be better described as automatic.</p>

References

- American National Standards Institute. *Processes for Engineering a System*, ANSI/EIA-632-1999. ANSI, 1999.
- Bailey, K. D. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks, CA: SAGE Publications, 1994.
- Clough, B. 'Metrics, Schmetrics! How the Heck Do You Aetermine a UAV's Autonomy Anyway?' In Proceedings of AIAA 1st Technical Conference and Workshop on Unmanned Aerospace Vehicles. Air Force Research Lab., Wright-Patterson AFB, OH, 2000.
- Dunlop. 'Interactions between Automatic and Autonomous Systems'. Presented at the Systems Engineering for Autonomous Systems Defence Technology Conference, July 2009, Edinburgh.
- Given, L. M. (ed.). *The SAGE Encyclopedia of Qualitative Research Methods* (Vol. 2). Thousand Oaks, CA: SAGE Publications, 2008.
- Institute of Electronic and Electrical Engineers (IEEE). *Standard for Application and Management of the Systems Engineering Process Description*. IEEE Std 1220-1998, 1998.
- International Organization for Standardization (ISO). *Robots and Robotic Devices* (ISO/TC184/SC2). Geneva: ISO, 2012. As cited in E. Prestes, et al. 'Towards a Core Ontology for Robotics and Automation'. *Robotics and Autonomous Systems* 61/11 (2013): 1193–1204. Available at <http://dx.doi.org/10.1016/j.robot.2013.04.005>, accessed 25 March 2015.
- Melzer, N. Human Rights Implications of the Usage of Drones and Unmanned Robots in Warfare. Brussels: European Parliament, Directorate-General for External Policies, 2013.
- National Aeronautical and Space Administration (NASA). *Systems Engineering Handbook*. NASA/SP-2007-6105 Rev1. Washington, DC: NASA, 2007.
- NATO Industrial Advisory Group (NIAG). *Pre-Feasibility Study on UAV Autonomous Operations*. NIAG Special Group 75, 2004.
- North Atlantic Treaty Organization (NATO). *NATO Glossary Of Terms And Definitions*. AAP-6, Edition 2014. Brussels: NATO Standardization Office, 2014.

- National Institute of Standards and Technology (NIST). *Autonomy Levels for Unmanned Systems (ALFUS) Framework: Volume 1—Terminology*, Version 2.0. NIST Special Publication 1011-I-2.0, 2012.
- Parasuraman, R., Sheridan, T.B., and Wickens, C.D. 'A Model for Types and Levels of Human Interaction with Automation'. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 30/3 (2000): 286–97.
- Proud, R. W., Hart, J.J., and Mrozinski, R.B. *Methods for Determining the Level of Autonomy to Design into a Human Spaceflight Vehicle: A Function Specific Approach*. Houston, TX: NASA Johnson Space Center, 2003.
- Register, The. 'Unmanned, autonomous ROBOT TRUCK CONVOY 'drives though town'. 3 February 2014. Available at http://www.theregister.co.uk/2014/02/03/unmanned_autonomous_robot_truck_convoy_drives_though_town/, accessed 25 March 2015.
- Scharre, P., and Horowitz, M. *An Introduction to Autonomy in Weapon Systems*. Working Paper – Project on Ethical Autonomy. Washington, DC: Center for a New American Security, 2015.
- Sheridan, T. B. *Telerobotics, Automation and Human Supervisory Control*. Cambridge, MA: MIT Press, 1992.
- Stensson, P. and Jansson, A. 'Autonomous Technology – Sources of Confusion: A Model for Explanation and Prediction of Conceptual Shifts'. *Ergonomics* 57/3 (2014): 455–70.
- Truszkowski, W., Hallock, H., Rouff, C., Karlin, J., Hinchey, M., Rash, J. L., and Sterritt, R. *Autonomous and Autonomic Systems: With Applications to NASA Intelligent Spacecraft Operations and Exploration Systems*. Berlin: Springer, 2009.
- UK Ministry of Defence (MOD). *The UK Approach to Unmanned Aircraft System: Joint Doctrine Note 2/11*, 2012.
- US Department of Defense (DOD). *Systems Engineering Fundamentals*. Fort Belvoir, VA: Defence Acquisition University Press, 2001.
- US Department of Defense (DOD). *Joint Publication 1-02, DOD Dictionary of Military and Associated Terms*, 8 November 2010.

US Department of Defense (DOD). (2012). *Task Force Report: The Role of Autonomy in DOD Systems*. Washington, DC: Defense Science Board, DOD: 2012.

Williams, A. P. *Typological Analysis: Autonomous Systems*. Report prepared for Multinational Capability Development Campaign, 2013-14. Norfolk, VA: NATO Headquarters Supreme Allied Commander Transformation, 2014.

Williams, R. *Autonomous Systems Overview*. BAE Systems, 2008.

PART 2:

ETHICAL, LEGAL, AND POLICY PERSPECTIVES

Developing Autonomous Systems in an Ethical Manner

Chris Mayer

Abstract

Military weapons and equipment benefit a great deal from advances in technology. These advances have allowed combatants to engage targets more accurately and from further away, which has meant increased safety for military forces and civilians and an improved capability to accomplish tasks. The latest technology to improve the capability of military systems is autonomy. Military systems are gaining an increased capability to perform assigned tasks autonomously, and this capability will only improve over time. Autonomy in military systems⁽¹⁾ allows human operators to remain out of harm's way and to complete tasks that manned systems cannot (or to complete them more efficiently). It also gives human operators more capability than if they used manned systems, as humans can direct multiple autonomous systems. If autonomous capabilities achieve the potential that many of its supporters believe it can, this will radically change how military operations are conducted.⁽²⁾ Because these changes may be so significant, and because they allow systems to perform tasks without being constrained or guided by constant human oversight or decisions, autonomous systems need to be examined from an ethical perspective. This is especially necessary because autonomy can have both positive and negative effects on human life. While many articles and chapters focus exclusively on either the negative or positive implications of autonomous systems, this chapter will consider both, and then offer recommendations for advancing policy in

1. Different names for systems with autonomous capabilities will be used for stylistic effect, but they all designate the same types of capabilities. These names also imply that these systems are not fully autonomous but only perform certain tasks autonomously. Essentially, an autonomous system is one that can perform certain functions autonomously.

2. For views on the benefits of autonomous weapons, see Arkin 2013.

a manner that takes into account ethical considerations for the development and use of autonomous systems.

Introduction

This chapter leaves aside the debate about what capabilities autonomous technologies will be able to achieve, and assumes that they will be able to perform according to the expectations of their supporters. If this turns out to be too optimistic, then there will be a greater burden on these supporters to argue why systems with greater and greater degrees of autonomy should be developed and deployed. While this chapter focuses solely on the ethical aspects of the use of autonomous systems, for the sake of the debate, supporters and opponents should come to an agreement on what the capabilities actually are, and what they will reasonably be in the future. This will make the dialogue about the ethics of these systems more productive. It is also not ethical to oppose or support technology such as autonomy without truly understanding both its benefits and drawbacks. Therefore, it is important to present an accurate portrayal of the capabilities that these technologies offer now – and might offer in the future.⁽³⁾

It is important to establish terminology and definitions when discussing autonomy, as ambiguity often leads to misdirected and incoherent arguments. This chapter uses the MCDC definition of autonomy: ‘Autonomous functioning refers to the ability of a system, platform, or software to complete a task without human intervention, using behaviours resulting from the interaction of computer programming with the external environment’.⁽⁴⁾ Debates about autonomous systems often focus on the assumption that these systems are fully autonomous, and that they possess a form of artificial intelligence that allows them to function in the same manner as a sentient being. The MCDC definition makes clear that such systems are not fully autonomous, and clearly distinguishes between systems with autonomous functioning and unmanned systems. This is an important distinction, as these two systems are often confused, yet autonomous systems offer benefits that unmanned systems do not.

3. I thank Paul Scharre for recommending that I include a brief discussion of this prior to the recommendations.

4. MCDC 2014, 9.

Military systems with autonomous capabilities may reduce the risk to military personnel by performing ‘dangerous, dull, and dirty’ tasks so that humans do not have to.⁽⁵⁾ This will keep humans out of harm’s way, and may accomplish the mission more effectively than if the task had been performed by a human or a manned system. While unmanned systems also have the ability to perform these tasks, autonomous systems can do so without human interaction. This provides greater capability when communication with a system is an issue, and minimises the number of human operators needed to accomplish particular tasks.

The use of systems with autonomous capability may also reduce unintentional harm to civilians during military operations if these systems are better able to discriminate between civilians and combatants due to an ability to make decisions faster and more accurately than humans. Vehicles that operate autonomously may have a much lower chance of error than human-operated vehicles due to the elimination of operator fatigue as a possible cause of an accident. The same is true for weapon systems that operate without humans ‘in the loop’, as operator fatigue can result in unintended harm, even when the system is operated remotely.⁽⁶⁾ This capability may also significantly improve the capabilities of military organisations by allowing them to deploy these systems without the need for constant human communication and control. This is especially beneficial where communication is eliminated or non-existent due to environmental conditions or enemy denial activities. These systems may also allow military operations to be accomplished at a reduced cost and with reduced personnel.

Despite the multitude of benefits promised by autonomous technology, there is also great ethical risk associated with their use. Over the past few years, numerous writers have argued against weaponising autonomous systems due to these associated ethical problems.⁽⁷⁾ Many of these arguments are practical, and relate to the systems’ capability and effectiveness. There are also purely ethical objections to their use. For example, Aaron Johnson and Sidney Axinn argue that ‘the decision to take a human life must be an inherently human

5. Diab 2014.

6. Autonomous systems with lethal capability (autonomous weapons) refer in this context to a weapon system that is able to locate and engage a target without a human ‘in the loop’.

7. See, for example, Sharkey 2010; Johnson and Axinn 2013; Vallor 2013; Lin 2010; HRW ND.

decision...it would be unethical to allow a machine to make such a critical choice.’⁽⁸⁾ This issue is addressed in greater depth below.

Thus several tensions and contradictions arise when considering the development and use of autonomous systems, especially autonomous weapons. In some ways, the systems’ increasing ability to perform certain functions autonomously offers the possibility to save lives and reduce the costs of military operations, thus supporting the value of human life. Yet there are many ethical concerns about their development and use. The rest of the chapter discusses these concerns in depth and offers guidelines on how to address this tension as policy makers consider the widespread use of autonomous technology for military systems.

How the Development and Use of Autonomous Systems May Enhance Respect for the Value of Human Life

One of the state’s primary obligations is to protect the lives of its citizens from each other and from citizens of other states.⁽⁹⁾ States fulfil this obligation by providing internal security in the form of police and other security organisations and external security in the form of military forces. A state’s obligation to protect lives also extends beyond its own citizens. When engaged in a conflict, a state’s military forces may only intentionally target military forces from another state, not civilians. Military forces may cause foreseen but unintentional harm to civilians (collateral damage), but this harm must be outweighed by the military value achieved by destroying the target, and the use of force must meet certain requirements. The guidelines used by the military to determine the permissibility of unintentional harm to civilians are known as the Doctrine of Double Effect (DDE), which has four conditions.⁽¹⁰⁾ The conditions prohibit intentional harm to civilians and require combatants to

8. Johnson and Axinn, 134.

9. Numerous political philosophers have proposed that states have this obligation. Thomas Hobbes and John Locke are two of the most cited sources of this view. Even a libertarian such as Robert Nozick, who advocates minimal government, argues that the state has the obligation to protect its citizens.

10. Michael Walzer describes the four conditions as follows: (1) ‘The act is good in itself or at least indifferent, which means, for our purposes, that it is a legitimate act of war; (2) The direct effect is morally acceptable – the destruction of military supplies, for example, or the killing of enemy soldiers; (3) The intention of the actor is good, that is, he aims only at the acceptable effect; the evil effect is not one of his ends, nor is it a means to his ends; (4) The good effect is sufficiently good to compensate for allowing the evil effect; it must be justifiable under Sidgwick’s proportionality rule.’ Walzer 2006, 153.

ensure that the military objective they seek to achieve by engaging the target outweighs any foreseen unintentional harm to civilians. Many theorists even argue that a state's military forces should seek to minimise unintentional harm to civilians when engaged in military operations. Michael Walzer advocates this view by seeking to revise the DDE's fourth criterion to suggest that military members should seek to minimise foreseen unintended harm, even if it requires them to assume increased risk.⁽¹¹⁾ Thus, states must respect the value of the lives of both their own citizens and those of other states by avoiding intentional harm and minimising unintentional harm.

Even though military members accept risk to their lives through service in the military, they still possess rights as humans, and the state owes them protection as it does all of its citizens. Although military members relinquish some of their rights, and many often find themselves in harm's way defending the interests of their state, it is reasonable for military members to expect the state to respect their rights. Brian Orend describes the state's obligation to military members as based on five fundamental human rights: security, subsistence, liberty, equality, and recognition.⁽¹²⁾ The most relevant right for this analysis is the right to security. Even though they face possible death by being asked to participate in combat, Orend argues that military members are entitled to: 'sound and serious military training; to be free from severe and dangerous inaugural or "hazing" rituals; and to good, functional equipment and weapons which enable them to perform their job.'⁽¹³⁾ He also asserts, 'Historical instances of deploying troops as mere "cannon-fodder", for instance, plausibly count as internal human rights violations.'⁽¹⁴⁾ Even though Orend is referring to poor leadership and training, his discussion of human rights violations and the value of military members' lives can also be applied to the use of autonomous systems to support or replace military personnel.

If a state can accomplish particular types of military missions by using systems with autonomous capabilities, while significantly reducing the risk to military members and without increasing the risk to civilians, then it is obliged

11. Ibid., 155.

12. Orend 2013, 133.

13. Ibid.

14. Ibid., 134.

to develop and use these systems, assuming that the cost is not prohibitive. To do otherwise would undermine the human rights of military members as well as the state's obligation to protect its citizens. For example, imagine that there is a requirement for the perimeter of a military base located in a combat zone to be patrolled, and that humans patrolling the base routinely face the risk of harm by enemy combatants. If an autonomous weapon system is able to patrol as effectively as human combatants and engage any intruders with lethal force, and if these systems can be produced at a reasonable cost, then a state arguably has an obligation to its military members to develop and use the system. Otherwise it exposes the military to unnecessary harm, thus failing in its obligation to respect the value of their lives.

Autonomous systems can also help reduce harm to civilians. Given that they will possess advanced sensors and will be able to handle increased amounts of information, autonomous systems may be able to make better targeting decisions or react more appropriately to unexpected events such as a child darting in front of an autonomously functioning supply truck. This advantage, combined with the fact that these systems will not have the human drive for self-preservation that may lead combatants to improperly aim weapons or directly harm civilians, will result in better discrimination between combatants and civilians and a reduction in accidental harm to civilians.

Another advantage is that while human error is often linked to fatigue, autonomous systems will not tire, and will thus avoid these sorts of errors that often result in harm to civilians and military forces as well as property damage. Autonomous systems can also be employed repeatedly without having to provide the rest time that human operators need. The more tasks that can be performed by autonomous systems, the less likely that human error caused by fatigue will cause harm.

Another way that the use of autonomous systems may reduce the risk to civilians is that the systems do not possess emotions or feel attachment to their comrades. Many war crimes are committed by combatants who seek revenge for the deaths of their fellow military members. Also, due to fear for their lives, or even for the destruction of an unmanned system, combatants might exercise less restraint than they should when engaging targets when civilians are near. This may lead to more unintended harm to civilians than is morally acceptable. For example, the operators of a manned supply truck may drive

recklessly due to a fear of being ambushed or captured. The use of autonomous systems reduces this possibility. As noted above regarding minimising harm to military members, if autonomous systems offer the capability to perform missions as effectively and as cheaply as manned systems, while measurably reducing unintended harm to civilians, then states have an obligation to develop and use them. To not do so would arguably devalue the lives of the civilians that autonomous systems could save, as well as the ethical principles that require one to minimise civilian harm.

In addition to reducing the loss of life (both military and civilian), cost should be taken into consideration, as military budgets often comprise a significant portion of states' budgets. As mentioned above, states have a responsibility to train and equip their military members well so that they can successfully accomplish assigned missions while minimising the loss of life. This has a limit, however, as states have other budgetary requirements. This means that states should aim to provide as effective a military force as possible in the most cost-effective manner as possible. While the cost of systems that can perform tasks autonomously varies significantly, for the sake of argument imagine that employing them for certain tasks will significantly reduce the costs of completing particular tasks. This would allow the savings to be used to better equip military members so that they can more effectively and safely accomplish their assigned tasks. Or, savings could be diverted to fund other state priorities, such as domestic programmes or foreign aid. If autonomous systems truly offer the same (or better) level of mission effectiveness at a lower cost, then states have an obligation to employ them. Not doing so wastes the resources given to them by their citizens.

If they fulfil the expectations that many have of them, the development and use of autonomous systems will allow states to affirm the value of human life by reducing potential harm to military members, civilians, and property. Autonomous systems may also allow military forces to complete assigned tasks more effectively and cheaply, which will let states shift funding to other requirements that, in some cases, may allow states to better fulfil other obligations they have to their citizens.

How the Development and Use of Autonomous Systems May Undermine Respect for the Value of Human Life

Despite the potential of autonomous systems to reduce the risk to humans, thus affirming the value of human life, there is also great ethical danger associated with advances in autonomy. Over the past few years numerous writers have argued against autonomous weapons, or ‘lethal autonomous weapons’ as they are often called, due to ethical concerns.⁽¹⁵⁾ Practical concerns focus on the limitations of these weapons, as there are questions regarding whether they will ever be able to accurately discriminate between civilians and combatants. Noel Sharkey argues that, ‘The ethical problem is that no autonomous robots or artificial intelligence systems have the necessary skills to discriminate between combatants and innocents.’⁽¹⁶⁾ Additionally, he believes that these types of systems would not be able to calculate proportionality, as ‘there is no sensing or computational capability that would allow a robot such a determination, and nor is there any known metric to objectively measure needless, superfluous or disproportionate suffering.’⁽¹⁷⁾ There are also concerns about the threat of hackers who could possibly gain control of these weapons and turn them against military forces or even civilians, which is also a problem for remotely operated systems. Finally, some wonder whether autonomous systems would be able to replicate humans’ moral reasoning when making targeting decisions. The most strenuous objection comes from those who argue that it is only ethically permissible for humans to intentionally harm other humans. They propose that autonomous systems should never be allowed to target and harm a human. There is also controversy related to using any autonomous weapon, even when it is not equipped with lethal capability.

This section will examine autonomy in tasks that are not related to the use of lethal force, as they raise issues that are far less controversial than autonomous weapons. In the future, there will most likely be supply vehicles that can operate autonomously. While these vehicles will not be armed with lethal capability, they could be taken over by an unauthorised user who could divert the vehicle into a crowd of civilians or their property. The vehicles could also

15. Three paradigmatic examples of such concern are presented in Sparrow 2007; Sharkey 2010; and Johnson and Axinn 2013.

16. Sharkey 2008, 87.

17. Sharkey 2008, 88.

be used against the military forces that own them, by either being caused to crash or directed against military members. This would seriously undermine the mission effectiveness of military units and hinder the ability of military forces to accomplish their mission. While this problem exists with many types of advanced technology, especially unmanned systems, autonomous systems will be less supervised by humans than other systems, which might allow autonomously driving vehicles to cause more damage than a remotely operated one, whose operator might be able to regain control or notice the loss of control earlier than someone monitoring an autonomous system.

The potential consequences if a military lost control of an autonomous weapons system is even greater. Autonomous weapons could be turned against civilians and military forces, and essentially be used to commit war crimes or to inflict significant casualties or equipment damage on military forces. For example, in 2007 a semi-autonomous weapon system used by the South African Army malfunctioned, killing nine soldiers and wounding 14 others.⁽¹⁸⁾ While this weapon was not hacked, the fact that it killed friendly forces highlights the danger of delegating the ability to fire a weapon from humans to an autonomous weapon. An hacked autonomous weapon system would be much more dangerous than the malfunctioning South African weapon, as it could be used to target multiple groups of people. Thus, opponents cite this potential harm to argue that autonomous weapons should not be developed or employed at all.

One possible solution is for autonomous systems to be connected to (and continuously monitored by) a human supervisor so that they can be shut down if necessary. Yet this would make the system even more vulnerable to hackers, and may negate one of the primary benefits of autonomous technology (reducing the need for personnel) and make the system more like a remotely operated system than an autonomous one.

Many sceptics suggest that it will be nearly impossible for an autonomous system to discriminate between civilians and combatants as well as humans, or that there should be a ban on autonomous technology until it can do so. This raises the question of how well an autonomous system should be able to discriminate before it is morally permissible to use it. At the very least, it seems

18. Shachtman 2007.

that an autonomous weapon must pose no more risk to civilians than human combatants currently do; ideally, they would pose less risk.

Some supporters of the use of autonomous weapons note that all weapons are legally reviewed prior to being placed in a state's arsenal. Thus, an autonomous weapon system that could not discriminate at all would be illegal and would not be used.⁽¹⁹⁾ Even if states adhered to this requirement, there is still concern about whether autonomous systems will ever be able to determine whether the military value of destroying a target is worth the unintentional but foreseen harm it would cause to civilians. Many believe that calculating whether the collateral damage caused by a military action is outweighed by the value of the target requires human judgement, and that autonomous systems will never have this capability or be able to follow a set of guidelines such as the DDE. While this heavily depends on how much the technology delivers, it is clear that autonomous weapons would need to be able to reliably calculate collateral damage costs, which is difficult even for humans to do.

Some worry about the increasing distance of the combatant from the battlefield, and the ease with which other combatants may be killed by weapons that allow targeting that is unmonitored by humans. Lamber Royakkers and Rinie van Est call those who fight from far away 'cubicle warriors.' They argue that the moral disengagement that takes place when using weapons that allow the enemy to be engaged from afar 'limits [the ability of] the cubicle warrior to reflect on his decisions and thus to become fully aware of the consequences of his decisions.'⁽²⁰⁾ This has been an issue as technology improves, especially with unmanned systems; however, it becomes even more of an issue with autonomous weapons, since humans would not be engaging the targets. Royakkers and van Est also claim that, 'Empirical evidence supports that moral disengagement leads to unethical behaviour.'⁽²¹⁾ The widespread use of autonomous weapons may further disengage combatants and policy makers from military actions, and they may fail to seriously consider the weapons' harmful effects on civilians and military members. An increasing disengagement, as autonomous weapons do more and more of the fighting, may in turn diminish combatants' and policy

19. Anderson, Reisner, and Waxman (2013, 386) articulate the conditions that must be met for a weapon to be considered lawful.

20. Royakkers and van Est 2010, 292.

21. Ibid.

makers' appreciation of the value of human life, especially for citizens of other states. This shift may eventually lower the bar for taking military action, which may make military conflict more frequent.

Military Virtues

Military conflict inevitably leads to the identification and celebration of heroes. Additionally, those who train to become members of their state's military forces develop particular virtues, such as courage and discipline, which are often celebrated and seen as paradigms of ideals. Many people join the military for the purpose of developing these virtues, and it is often noted how a young person who completes military training 'becomes a new person'. This attitude is also important for the sustenance of a state and the recognition of the military's value. While only a small percentage of the population of modern democracies serves, the rest of the population celebrates these few for their service. They serve as representatives of heroism, and the state's top leaders award its top military awards. Widespread use of autonomous systems, however, may minimise the opportunity for (and value of) military service.

This has been a continuing worry as technology has enabled combatants to fight from a distance. However, if wars were truly fought using a large number of systems with autonomous functions, especially autonomous weapons, then it is quite possible that this conception would change. Shannon Vallor argues that 'a well-designed and implemented robot can promote the interest, and specifically the security of, a nation. Yet it is not clear how a robot could "have" a country to serve.'⁽²²⁾ This commitment of service to a state, as articulated by the oath of office taken by military members, emphasises the importance of the state and its people. Vallor suggests that 'the development and deployment of armed military robots on a wide scale may problematise the modern military ideal of selfless service, and may displace the traditional virtues of military character that have historically been cultivated within the armed forces to promote the effective realization of that ideal.'⁽²³⁾ Vallor also proposes that, 'ideals of military virtue also play a significant role in how soldiers are perceived by the publics they serve, and by foreign combatants and

22. Vallor 2014, 173.

23. Ibid.

civilians.²⁴) Even apart from undermining conceptions of service, the use of autonomous systems for dangerous, dull, and dirty tasks may leave military forces unprepared to perform these tasks themselves should the need arise, or lead to the perception that these sorts of tasks are too risky to be performed by people. Thus the increased use of autonomous weapons may, paradoxically, improve the security of the state while at the same time put it on shaky ground, since it may leave the state's military forces unprepared to fight a war without autonomous weapons. Thus, the development and use of autonomous systems has the potential to undermine the military's professional ethic in a number of important ways that may harm the state.²⁵

Another risk of using a great number of autonomous systems is that it reduces the risk to military forces. While this is good news for members of the military forces using the autonomous systems, there is the possibility that enemy combatants will become frustrated at their inability to inflict casualties on those forces. The types of states using autonomous systems are very sensitive to military casualties, and seek to minimise them in order to avoid losing public support for military operations. Enemy combatants know this, and seek to inflict casualties on those they fight. Given that destroying autonomous systems will not have the same impact as causing casualties, enemy combatants may redirect their efforts toward causing civilian casualties.

It is of course true that there are all sorts of technological advances that help military forces avoid or reduce casualties. What is different about autonomy, however, is that it has the potential to remove a significant number of forces from the battlefield, allowing states to maintain support for conflicts much longer than they do now and making it much more difficult for enemy forces to kill combatants. If (and this is most likely decades away) it ever becomes possible for autonomous weapons to take and hold ground, it will make inflicting casualties by enemies extremely difficult (unless, of course, the autonomous

24. Ibid.

25. A question that arises when considering this issue is whether autonomous weapons are weapons in the traditional sense (like a rifle or a plane), or whether they are combatants in their own right. The worry is that they become more and more like combatants as they independently perform tasks rather than having to rely upon humans. Combatants fire rifles. With advanced missiles, combatants select targets and then fire the missile. The more that autonomous weapons appear to be making decisions, the less they appear to be weapons and the more they are combatants in their own right. I thank Paul Scharre for raising this important point, and for highlighting the fact that only a small number of citizens serve in volunteer armies.

weapons are so expensive that their destruction is able to persuade a populace to remove its forces from a conflict). Thus, by employing autonomous systems, military forces may be deflecting harm away from themselves to civilians. This is ethically problematic, as military forces are meant to protect civilians rather than endanger them.

The most significant ethical issue regarding the value of human life concerns whether it is ethically permissible for an autonomous weapon to take the life of a human. Even though there is disagreement about the philosophical foundations of this debate, philosophers generally agree that humans have value and that there have to be good reasons to take a human life. Humans' ability to reason and to develop a life plan is seen as special and, generally, it is only permissible to intentionally take the life of a human in self-defence. Thus, allowing autonomous weapons to 'decide' to target a person is morally impermissible, as the autonomous weapon does not have the right to self-defence. What is meant by 'decide' here is multifaceted, and comes at different stages of an autonomous weapon's use. The first aspect concerns an autonomous weapon searching for a set of possible targets. Another aspect concerns the autonomous weapon selecting a target from the set of possible targets. This aspect has numerous parts, which include deciding which objects are legitimate targets from the set of possible targets, and then determining whether it is permissible to engage the legitimate targets (e.g., perhaps there are nearby civilians that raise proportionality concerns). The final aspect concerns the rules of engagement to use. This is a decision that an autonomous weapon might make if it has numerous possibilities from which to select. As it moves from an open area to a town, it will have to select appropriate rules of engagement. All of these aspects of an autonomous weapon's 'decisions' have ethical implications.

Additionally, human life is so valuable because only humans possess both rationality and emotions and are capable of deciding whether it is worth taking a life. For opponents of autonomous weapons, giving autonomous systems the latitude to target and kill (at their discretion) combatants is ethically problematic. This is especially true given the responsibility that the state gives to members of the military. Societies train military members to make the proper decision regarding whether to kill. If autonomous weapons are truly making decisions, then it is they (or, at least, the developers who made them or the

programmers who programmed them) that are making the decisions to kill, which undermines the military professional ethic.⁽²⁶⁾

Additionally, many ground the moral permissibility of intentionally harming a human in the right to self-defence (i.e., it is only permissible to intentionally harm another human when that person is performing an action that will likely harm you). Many believe that all combatants have the right to self-defence, while others believe that only those on the just side have the right to intentionally harm during war. Regardless, however, it is the person's value that permits him or her to intentionally harm in the name of self-defence. Even if the combatant's right to intentionally harm is due to the fact that he or she acts as an agent of the state, there is something special about the combatant that allows him or her to serve as an agent of the state. The military member is both a citizen of the state and a representative/protector of the state's populace. As Vallor's comments above suggest, it is difficult to imagine how anything other than a human could act as an agent of the state. Without the right to self-defence or the ability to serve as an agent, it is difficult to see how it could be morally permissible for an autonomous weapon to intentionally harm a human.

Those who do not see ethically prohibitive reasons against autonomous weapons targeting combatants address what they see as weaknesses in their opponents' argument. For example, they claim that autonomous weapons do not actually decide to target and harm combatants. Rather, they are directed to certain areas and operate within mission parameters. Humans decide to place autonomous weapons in certain areas, and humans program autonomous weapons to target people who meet certain criteria. The people who meet these criteria are deemed combatants and thus may be targeted. While autonomy might someday reach the point at which autonomous weapons can truly act independently, supporters would claim that the technology is not even close to the point where this is possible. Additionally, supporters claim that autonomous systems are not agents but merely weapons. While they have a different capability than other currently available systems, they should be treated the same as other systems (i.e., those who direct them to engage targets are responsible for their actions and can be held responsible).

26. I thank Paul Scharre for raising this concern during his review of my essay.

Conclusion

As the section above makes clear, there are numerous ethical concerns associated with the development and use of autonomous systems. Concerns over their technological effectiveness and their ability to respect human life in the same manner that other humans can is a common theme, as is the danger that these systems will be misused or cause harm due to their autonomy. However, the biggest worry relates to placing the decision to take a human life in the hands of a machine.

Recommendations

Below are recommendations for policy makers that will allow ethical considerations to play a proper role in developing and using autonomous systems.

1. Seek agreement on the advantages offered by technological improvements in autonomy as well as the financial savings (if any) that would result from incorporating increasingly autonomous functions into military systems. Proponents and opponents of autonomous capability in military systems should determine what can reasonably be expected from advances in autonomous capability. Overly pessimistic or optimistic views on autonomous capabilities lead to distrust and, more importantly, the inability to accurately determine the ethical impact of developing and using autonomous systems during military operations.
2. When possible, be as transparent as possible about the results from the testing and use of autonomous systems. This will make achieving the first recommendation easier.
3. Both supporters and opponents of autonomous technology should define what they mean by autonomy, artificial intelligence, agency, intention, and other key terms as they relate to the development and use of autonomous systems for military operations. For example, there are significant ethical implications associated with the difference between autonomy, as it is commonly understood, and artificial intelligence. Common definitions will make the dialogue between supporters and opponents much more productive.
4. Explain whether autonomous systems actually ‘decide’ to engage a target, and how they are different (if at all) from other systems that allow

combatants to engage enemies from a great distance using different aspects of an autonomous weapon's 'decision' as presented above. This will establish the necessary foundation for a conversation that considers whether it is inherently wrong for an autonomous system to 'intentionally' harm a human.

5. Engage in a discussion on how, if at all, military member responsibility for accidental harm or property damage differs when using autonomous systems as compared to other military systems. This is relevant for autonomous systems that perform both lethal and non-lethal tasks.
6. As the ethical benefits and risks associated with using autonomous systems become more apparent, policy makers should identify these facts and be as transparent as possible about any trade-offs involved in using autonomous systems.

References

- Anderson, Kenneth, Reisner, Daniel, and Waxman, Matthew. 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems'. *International law Studies* 90 (2014): 386–411.
- Arkin, Ron. 'Lethal Autonomous Systems and the Plight of the Non-combatant'. *Artificial Intelligence and Simulation of Behavior Quarterly* 137 (2013): 1–9.
- Diab, Ann. 'Drones Perform the Dull, Dirty, or Dangerous Work'. *Tech Cocktail*, 12 November 2014. Available at <http://tech.co/drones-dull-dirty-dangerous-2014-11>, accessed 20 November 2014.
- Human Rights Watch (HRW). 'Killer Robots', ND. Available at <http://www.hrw.org/topic/arms/killer-robots>, accessed 25 March 2015.
- Johnson, Aaron M. and Axinn, Sidney. 'The Morality of Autonomous Robots'. *Journal of Military Ethics* 12/2 (2013): 129–41.
- Lin, Patrick. 'Ethical Blowback from Emerging Technologies'. *Journal of Military Ethics* 9/4 (2010): 313–31.
- Multinational Capability Development Campaign (MCDC). *Policy Guidance: Autonomy in Systems*. Norfolk, VA: Supreme Allied Commander Transformation HQ, 2014.
- Orend, Brian. *The Morality of War*, 2nd edition. Ontario: Broadview Press, 2013.
- Royakkers, Lamber, and van Est, Rinie. 'The Cubicle Warrior: The Marionette of Digitalized Warfare'. *Ethics in Technology* 12 (2010): 289–96.
- Shachtman, Noah. 'Robot Cannon Kills 9, Wounds 14'. *Wired*, 18 October 2007. Available at <http://www.wired.com/2007/10/robot-cannon-ki/>, accessed 16 March 2014.
- Sharkey, Noel. 'Grounds for Discrimination: Autonomous Robot Weapons'. *RUSI Defence Systems* (October 2008): 86–9.
- Sharkey, Noel. 'Saying "No!" to Lethal Autonomous Targeting'. *Journal of Military Ethics* 9/4 (2010): 369–83.
- Sparrow, Robert. 'Killer Robots'. *Journal of Applied Philosophy* 24/1 (2007): 62–77.
- Vallor, Shannon. 'The Future of Military Virtue: Autonomous Systems and the Moral Deskillling of the Military'. Paper presented at the 5th International Conference on Cyber Conflict, 4–7 June 2013, Tallinn, Estonia. Available

at https://ccdcoe.org/cycon/2013/proceedings/d2r1s10_vallor.pdf,
accessed 25 March 2015.

Vallor, Shannon. 'Armed Robots and Military Virtue.' In *The Ethics of Information Warfare*, edited by Luciano Floridi and Mariarosaria Taddeo, 169–85. New York: Springer International Publishing, 2014.

Walzer, Michael. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books, 2006.

The Legal Implications of the Use of Systems With Autonomous Capabilities in Military Operations

Roberta Arnold⁽¹⁾

Abstract

Developments in the field of technology have led to an increased use of so-called systems with autonomous capabilities (SAC) for both military and civilian purposes. Notwithstanding their increasingly common presence in everyday life, their functioning and way of operating remains a question mark for many. The “mystery” of autonomous systems—associated by some with futuristic views of a world led by robots and androids replacing human beings—has been, at the same time, the source of legal concerns among some circles, especially in relation to their use within the military context. Several initiatives have been launched to address some of these concerns, including the 2013-15 autonomous systems research project led by NATO Allied Command Transformation under the Multinational Capability Development Campaign (MCDC) collaboration. A study group composed of lawyers, including the author, was set up to consider the legal implications of the use of such systems by the military under international law, including e.g. the Laws of Armed Conflict and International Human Rights Law. This chapter illustrates the outcomes of the research with regard to the application of the law of state responsibility, which aimed at identifying and examining the principles according to which responsibility for wrongdoings under international law committed by or through the use of SAC may be attributed to a state. The main conclusion is that there currently appears to be no need to adopt a SAC-specific treaty in order to address concerns related to the attribution of state responsibility for

1. The views expressed here are the author's only and, in particular, do not necessarily reflect those of the Swiss Federal Department of Defence. Particular thanks go to the other colleagues of the legal group within the MCDC studies for the comments made to previous drafts of the present chapter and to Paul Scharre at the Center for a New American Security for his useful remarks.

the (mis-)use of such systems. Due to their increasing complexity, however, and the fact that it will become more difficult for the user to predict their behaviour in complex environments, the stage of the technological development, and the associated risks and implications should be monitored by states and taken into account when testing new systems.

Introduction

Aim and Objective

Developments in the field of technology have led to an increased use of drones and other similar systems and tools, for both military and civilian purposes. They have become part of people's daily life, and have even been featured in an off-Broadway play.⁽²⁾ But even though they have been integrated into everyday language, their functioning and their way of operating remains a question mark for many. Major confusion particularly revolves around the notion of 'autonomous systems', a term that is often interchanged with expressions like 'remotely controlled systems', 'unmanned aerial vehicles', and 'robots'. At the same time, some circles have expressed growing legal concerns about their use by the military.⁽³⁾ The military and the International Committee of the Red Cross (ICRC)⁽⁴⁾ have not remained passive; they have joined several initiatives to address some of these concerns.

This chapter illustrates some of these concerns within the framework of the Multinational Capability Development Campaign (MCDC) study group on autonomous systems. The use of systems with autonomous capabilities (SAC)⁽⁵⁾ for *military* purposes⁽⁶⁾ has several legal implications, most notably under the *jus in bello* (in particular the Laws of Armed Conflict (LOAC) and International Human Rights Law (IHRL)), the *jus ad bellum* (in particular

2. A new off-Broadway show dedicated to drones and unmanned aerial vehicles is *Grounded* by George Brant. See Nudd 2015.

3. For example, see the Campaign to Stop Killer Robots, <http://www.stopkillerrobots.org/>.

4. ICRC 2014.

5. This chapter uses the term SAC to refer to machines that can assess a situation and choose a particular course of action independently of human supervision, but where human beings are nevertheless always somewhere in the loop.

6. The use of such systems for other purposes was beyond the scope of the MCDC study, which focused on the legal implications of the use of AxS for purely military operations.

the UN Charter and the law of neutrality), the law of state responsibility, and International Criminal Law (ICL). The MCDC study aimed to provide policy guidance on a number of selected international legal issues that appeared to be particularly relevant for the current use of autonomous systems:⁽⁷⁾

1. legality of weaponised SAC;
2. legality of the use of systems with autonomous weapon functions in combat (targeting law);⁽⁸⁾
3. legality of employing SAC for law enforcement and self-defence;
4. state responsibility for harm caused by SAC; and
5. criminal responsibility for harm caused by SAC.

It is beyond the scope of this chapter to venture into all these issues. Its focus will be on the legal implications of the use of SAC under the law of state responsibility, which establishes the principles according to which responsibility for wrongdoings under international law may be attributed to a state.

Yet it is nevertheless important to keep the larger legal picture in mind, and the fact that other legal regimes play an important role. An exact assessment of the legal implications of the use of SAC for military operations requires a comprehensive approach. Moreover, the shortcomings of one legal regime in addressing the problems raised with regard to the use of SAC may be balanced by solutions proposed by other legal regimes.

According to the law of state responsibility, states purchase military (weapon or weapon carrier) systems for use by their own armed forces. States also sign international conventions and treaties regulating the conduct of hostilities and respect for human rights, thus they are the ones that ultimately have to respond to breaches of these obligations vis-à-vis other states and their own citizens. Thus, it is important for them to know which SAC may be developed, purchased, and deployed in order to be compliant with their legal obligations.

The following subsection briefly illustrates the applicable international legal regime, the relevance of state responsibility in the overall applicable legal context, and its relationship to the other relevant areas of international law in

7. The details can be found in MCDC 2014, 14. The author would like to thank the other colleagues of the study group for their support.

8. These are systems in which the autonomous capability selects and engages targets on its own.

this regard. The next section provides more detail about the principles of state responsibility for the use of SAC, and is followed by the conclusions.

The Applicable International Legal Regime

For purposes of clarity and simplicity, it will be argued that two major sets of rules can be identified under international law: (1) rules applying in time of peace and (2) rules applying in times of armed conflict. In the past, military operations predominantly fell into the second category, but nowadays most armed forces also tend to be deployed in situations that do not qualify as an armed conflict under the *jus in bello* (e.g., to support civilian law enforcement authorities or to intervene in case of catastrophes). Thus the MCDC research group agreed to take a broad approach and to consider the various implications (legal, ethical, operational, etc.) within the various contexts.

From a legal perspective, however, the major divide between the applicable rules is dictated by the existence of an armed conflict. A situation of armed violence must meet the criteria of the definition of armed conflict pursuant to the LOAC (in particular articles 2 and 3 common to the Geneva Conventions of 1949) in order to be subject to the special rules applicable to warfare. This distinction is important, because only in this case can the members of a regular armed group claim combatant status (and, thus, immunity) for the conduct of hostilities. In consideration of the necessities of war, members of a regular armed force are allowed to kill the enemy as long as their actions conform with the four core principles of the LOAC: distinction, military necessity, proportionality, and limitation.⁽⁹⁾

In times of peace, IHRL applies instead. In times of peace, for instance, the right to life is much stronger than in times of armed conflict; thus military forces deployed to fulfil a law enforcement mission (as opposed to a warfare mission) have different rules of engagement. Under the LOAC, there are special rules requiring state parties to the Additional Protocol (AP) of 1977 to the

9. The principle of limitation determines permitted and prohibited means. For further details, see Kolb 2015, 80: 'The principle of limitation negates the admissibility of total war, where a belligerent could do everything it sees fit in order to overpower the enemy. ... The principle of limitation operated as a break on the principle of permissive military necessity.' See also Kalshoven 2007; Gomez 1998.

four Geneva Conventions of 1949 to review the legality of all new weapons.⁽¹⁰⁾ The question, thus, is whether the use of SAC raises unique issues under the different legal regimes applicable to different types of military operations. Care must be taken not to confuse legal issues that may arise from the complexity of international law, regardless of whether the means used by the military is autonomous or not. For instance, with regard to the use of so-called armed SAC within a law enforcement context, drawing up a list of (human) targets that should be identified and hit by a machine raises questions that do not relate to the type of operator (machine or human being), but rather to the conformity of the operation with human rights standards (e.g., the prohibition of extrajudicial killings and the principle of presumption of innocence). A different problem, then, is the use of ‘killer robots’ in the sense of systems that are ‘autonomous weapons’ in the choice of the (human) targets. Even in this case, however, it would be wrong to assume that a SAC would be fully autonomous, since in this case the targets would be chosen based on specific parameters that are pre-determined and pre-programmed by a human ‘in the loop’. With these distinctions in mind, the following branches of international law are the most relevant for the military use of SAC: the LOAC, IHRL, ICL, the *jus ad bellum*, and the principles regulating state responsibility for breaches of international obligations.

State Responsibility under International Law

Acts contrary to international law engage the international responsibility of the state to which they are legally attributable. One major concern arising in relation to the military use of SAC is the apparent uncertainty regarding the attribution of potential harm caused by them in contravention of the LOAC, IHRL, or the *jus ad bellum* (e.g., UN Charter, art. 2). State responsibility may be an issue, for example, when SAC are resorted to in self-defence in cross-border operations, when the geographical boundaries of the battlefield have been ‘stretched’ to a third neutral country.⁽¹¹⁾ For example, State A and State B are

10. For instance, AP I, art. 36 states that: ‘In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.’

11. See Heyns 2013, §80.

at war with each other. Members of the armed forces of B have infiltrated the territory of (neutral) State C, to conduct operations against State A. State A, in response, attacks the military forces of State B located in State C through the use of SAC. In this scenario, one may argue that the 'geographic boundaries' of the armed conflict between States A and B have been 'stretched' over to State C, even though C is not a party to the conflict. In this case, the major problem is the choice of the applicable legal regime (*jus in bello* or law applicable in peacetime) and the fact that State C may invoke a violation of its sovereignty by the use of SAC across its borders.

Another conceivable concern is that state attribution for the misconduct of SAC may be impaired by the reduced/remote human involvement. However, SAC are not synonymous with unmanned (i.e., remotely controlled) systems. A remotely controlled (unmanned) system always has an operator, even though he or she is not on board the vehicle. A SAC, by contrast, not only does not have an operator on board, but it is also capable of exercising some functions autonomously. In both cases, however, there will be a human involved, albeit at different stages of the chain. In the first case, the operator will be off-board somewhere, remotely directing the machine. In the second case, even though SAC may autonomously assess a situation and choose a particular course of action independently, their programming and deployment will nevertheless be based on decisions taken by human state agents whose conduct is attributable to the operating state.⁽¹²⁾

Different legal regimes, such as individual civil and criminal liability and state responsibility, should not be confused. The law of state responsibility establishes criteria for attributing conduct to states and regulates the legal consequences of this attribution, including the duty to terminate unlawful conduct and to provide reparation for the resulting harm. Unlike *individual* criminal responsibility, *state* responsibility is not based on the notion of personal culpability, but on the attribution to the state of the (mis-)conduct of its organs/agents.

This analysis does not seek to specify what international legal obligations may be breached by resorting to SAC, but whether the *general principles* for determining state responsibility – which are primarily codified in the

12. See, e.g., Earle 2013-14, § 32; MOD 2013-14.

International Law Commission's 2001 *Articles on Responsibility of States for Internationally Wrongful Acts* (ARSIWA)⁽¹³⁾ – are sufficient to determine state liability for the military use of SAC resulting in breaches of the law and harm.

Analysis of the Applicability of the Principles of State Responsibility for the Use of SAC

General Principles of State Attribution

Pursuant to ARSIWA, art. 5, a breach of international law by a state entails its international responsibility. An internationally wrongful act may consist of one or more actions or omissions or a combination of both. Whether there has been an internationally wrongful act depends on the requirements of the obligation and the framework conditions for such an act.⁽¹⁴⁾ Once a breach has been determined, the (mis-)conduct may then be attributed to a state when this was exercised by one of its *organs*, i.e., a state *body* exercising legislative, executive, judicial, or any other functions. State organs are defined as such by the internal law of the concerned state (ARSIWA, art. 4). Moreover, the conduct of persons or *entities* empowered by a state to exercise elements of governmental authority shall also be considered an act of the state, provided the person or entity is acting in that capacity in the particular instance (ARSIWA, art. 5). At the same time, the state is also held accountable for 'acts committed by persons forming part of its armed forces',⁽¹⁵⁾ regardless of whether these were acting in their official capacity.

Major current debates revolve around the misuse or malfunctioning of autonomous weapon systems (AWS), which are a sub-category of SAC; once activated, such systems select and engage targets on their own without further human approval. However, the legal issues that may arise for the purpose of state attribution are not AWS specific. The latter should be considered as any

13. See UN General Assembly 2001, 43. The International Law Commission (ILC) is a UN commission established in 1947 that promotes the progressive development of international law and its codification. UN General Assembly Resolution 56/83 of 12 December 2001 annexed the ARSIWA and commended them to the attention of governments. The ARSIWA are the successors of the ILC's Draft Articles on State Responsibility, and consist of the codification of the principles of the law of state responsibility.

14. ILC 2001.

15. Hague Regulations of 1907, art. 3 and AP I, art. 3. See also Melzer 2013, 38; Sassoli et al. 2012, 463.

other military weapon system, regardless of the degree of autonomy. Moreover, also in this case, the AWS' use by the *armed forces* of a state results in the latter's automatic responsibility.

Should the SAC (or AWS) have reached such a degree of autonomy to be able to exceed its authority or 'disobey' the instructions imparted by its superior or programmer, for example in the event of their unintentional malfunctioning, the state nonetheless remains liable for the activity of its organ or empowered entity (ARSIWA, art. 7).

Other major concerns in public opinion are related to the use of AWS by non-state actors, including criminal gangs. Since these cannot be considered state agents or organs, harm resulting from their use of AWS cannot be attributed to any state. This, however, does not prevent the victims from seeking redress via alternative judicial remedies (e.g., under criminal law or torts law, which provide for the individual responsibility of the perpetrator, which may be the manufacturer, programmer, or commander, depending on their employment). The law should always be looked at comprehensively: claims that may not be brought forward under the state responsibility doctrine may be admissible under different legal paradigms.

Circumstances Excluding Wrongfulness of an International Obligation

ARSIWA notes that:

[A]ttribution must be clearly distinguished from the characterization of conduct as internationally wrongful. Its concern is to establish that there is an act of the State for the purposes of responsibility. To show that conduct is attributable to the State says nothing, as such, about the legality or otherwise of that conduct, and rules of attribution should not be formulated in terms which imply otherwise.⁽¹⁶⁾

Thus, the issue at stake is not whether the underlying conduct was wrongful, but rather whether it can be attributed to a state, and whether it may lead to compensation and redress in case of wrongfulness. This is an important

16. ILC 2001, 39.

difference, for example with regard to ICL, for the purposes of attribution to a specific individual person.

This does not mean, however, that the state has an absolute, abstract responsibility, with no possibility to resort to defence. The ARSIWA also provides for defence arguments, excluding the attribution of responsibility to the state.

However, pursuant to ARSIWA, art. 27, the invocation of a circumstance *precluding* wrongfulness (see ARSWIA art. 2: when conduct consisting of an action or omission is attributable to a state under international law and constitutes a breach of an international obligation of the state), is *without prejudice* to compliance with the obligation in question (if and to the extent that the circumstance precluding wrongfulness no longer exists), nor to the question of *compensation* for any material loss caused by the act in question. This means that even in the event that a State may rightly invoke a circumstance precluding wrongfulness under ARSIWA, the state may nonetheless be expected to comply with the obligation that was breached and to make good any material loss suffered by another State due to this breach.⁽¹⁷⁾ This is another important difference from ICL, pursuant to which compensation will depend on wrongfulness (in addition to culpability, which is a concept foreign to state responsibility). The wrongfulness of an international obligation may be excluded in the following cases (defence arguments):

- if there is a valid consent by the state affected by the wrongful act (ARSIWA, art. 20);
- self-defence (ARSIWA, art. 21);
- if countermeasures were taken in respect of an internationally wrongful act (ARSIWA, art. 22);
- *force majeure* (ARSIWA, art. 23);
- distress (ARSIWA, art. 24);
- necessity (ARSIWA, art. 25); and
- compliance with peremptory norms (ARSIWA, art. 26).

Consent and self-defence

ARSIWA, art. 20 states that: ‘valid consent by a State to the commission of a given act by another state precludes the wrongfulness of that act in relation

17. Ibid., § 4; ICJ 1997, 39.

to the former state to the extent that the act remains within the limits of that consent.

According to some, where State B agrees to State A's use of SAC within its territory, e.g., to eradicate hostile elements (targeted killings scenario),⁽¹⁸⁾ State B's lack of knowledge of the SAC's functioning may impair its consent. A counterargument is that in such an agreement not only the mission, but also further aspects like responsibility in case of failure, would have to be discussed in advance and agreed upon. State B could restrict its consent to a specific, well-defined, type of intervention. If, for example, the mission is the killing of targets X and Y, States A and B would have to define the limits of the intervention and the conduct to be adopted if a breach of the law occurs. As already mentioned, however, attribution of responsibility to State A for failing to respect these limits would arise regardless of the type of weapon used.

Force majeure

A situation of *force majeure* is characterised by the following cumulative elements:⁽¹⁹⁾ (1) the act must be brought about by an irresistible force or an unforeseen event (2) which is beyond the control of the state concerned, and (3) which makes it materially impossible in the circumstances to perform the obligation.

The *material impossibility* to comply with an international obligation may be due to a natural or physical event (e.g., stress of weather that may divert state aircraft into the territory of another state, earthquake, floods, or drought), human intervention, or both. The state concerned must have no real possibility of escaping the effects of the exceptional event.⁽²⁰⁾ The difficulty of attributing state responsibility under such circumstances is not SAC-specific. It is important to note, however, that in the event of accidental, unintended, or undesired injury due to the negligent or reckless conduct of the state, *force majeure* cannot be invoked.⁽²¹⁾

18. Heyns, § 82.

19. ILC 2001, 76, § 2.

20. Ibid., § 3.

21. Ibid., §§ 2–3 and 9–10; Melzer 2013, 39.

In the event of a malfunctioning system, the fault may lie at different levels of the chain (e.g., with the manufacturer of the software, or with the testing system adopted by the state). If those responsible acted as organs or agents of a state, the injured state may invoke state responsibility under ARSIWA, art. 4. The injured individuals may also seek redress under the legal regimes providing for individual responsibility (e.g., criminal or torts law), even if those responsible were acting on behalf of the state (see ARSIWA, art. 58).

In the alternative scenario that the manufacturer of the malfunctioning system was a third state, once again the injured state may seek redress under the doctrine of state responsibility, pursuant to ARSIWA, art. 42.⁽²²⁾

Reparation and Compensation

As already discussed, under ARSIWA, art. 27, the invocation of one of the circumstances precluding wrongfulness is without prejudice to the question of compensation for any material loss caused by the act. It will be for the state invoking such a circumstance to agree with any affected state on the possibility and extent of the compensation payable in a given case. This option of a specific agreement on compensation is to be welcomed from a pragmatic point of view. In most cases, it may prove difficult to pinpoint the exact fault. This obstacle may be overcome by solving the matter with an agreement.

More detailed rules on reparation and compensation may always be foreseen in specific treaties. ARSIWA, art. 55 confirms that the ARSIWA rules do not affect the application of *lex specialis*.⁽²³⁾ For instance, Hague Convention IV⁽²⁴⁾ (HC IV), art. 3 and AP I, art. 91,⁽²⁵⁾ which belong to the special regime of the LOAC, provide that in the case of violations of the LOAC by a state's

22. According to this, a state is entitled to invoke the responsibility of another state if the obligation breached is due to (a) that state individually or (b) a group of states including that state. Additionally, the issue may be one of individual, manufacturer responsibility.

23. Like the LOAC rules stating that a state shall be responsible for the conduct of its armed forces, under HC IV Hague Convention, art. 3 and AP I, art. 91.

24. IV HC, art. 3: 'A belligerent party which violates the provisions of the said Regulations shall, if the case demands, be liable to pay compensation. It shall be responsible for all acts committed by persons forming part of its armed forces.'

25. AP I, art. 91: 'A Party to the conflict which violates the provisions of the Conventions or of this Protocol shall, if the case demands, be liable to pay compensation. It shall be responsible for all acts committed by persons forming part of its armed forces.'

armed forces, that state is bound to pay compensation to the injured state. This principle now forms part of customary law. In the event of breaches of IHRL, a similar obligation is foreseen by the International Covenant on Civil and Political Rights, art. 2, African Commission on Human and People's Rights, art. 7, , art. 25, and European Convention on Human Rights, art. 13.

In addition to redress under the doctrine of state responsibility (which may only be invoked by another state), breaches of the law that cause harm to individual victims may be redressed under other legal regimes such as ICL (International Criminal Court Statute, art. 75).⁽²⁶⁾ The right to seek compensation under HC IV, art. 3 and API, art. 91 is restricted to injured states. In the event of breaches of the LOAC within the framework of a non-international armed conflict, since the victims are the nationals of the state responsible for the violations, the lack of redress under the regime of state responsibility is compensated for by remedies provided under IHRL.⁽²⁷⁾

Although these problems are not SAC-specific, it is important to address them in order to highlight the fact that some concerns raised in relation to their use will eventually be addressed in general, and that they should not be used to imply that SAC are intrinsically problematic.

Conclusions

The principles of general international law governing state responsibility appear to adequately regulate the international responsibility and consequences of harmful acts resulting from the military use of SAC. Where the use of SAC is decided by a *de jure* or *de facto* state agent or organ, harmful use deriving from breaches of international obligations (e.g., under the LOAC, IHRL, or the UN Charter) may be attributed to a state.⁽²⁸⁾ The same holds especially true if the harm was the result of the SAC's (mis-)use by a state's armed forces. In this case, the state will be automatically accountable, due to the special state responsibility provisions foreseen by HC IV, art. 3 and AP I, art. 91. Most importantly, the concern that states may evade international responsibility

26. Melzer 2013; 41–2.

27. See Sassoli et al. 2012, 462–3.

28. The same conclusions were drawn from Heyns 2013, § 104.

based on their unawareness of the system's faults is mitigated by the fact that, in cases of negligence and/or reckless conduct, *force majeure* cannot be invoked.

In practice, the main problem will not be the attribution of legal responsibility, but the availability of effective judicial remedies for injured states and individuals, and in the victim's choice of the most appropriate judicial remedies. Claims under the law of state responsibility may be only brought forward by another state. The same holds true for compensation claims concerning breaches of the LOAC. Individual victims will instead have to seek redress under other legal regimes (e.g., IHRL or ICL). This problem, however, is not SAC-specific and must therefore be addressed and resolved on a more general level.

Thus, there currently appears to be no need to adopt a SAC-specific treaty to address concerns related to the attribution of state responsibility for the malfunctioning or misconduct of such systems. Concerns about the 'accountability gap' in general refer to legal aspects associated with *individual* accountability, not *state* responsibility. It is possible that as autonomous systems become increasingly complex, it will become more difficult for the user to predict their behaviour in complex, real-world environments, which could lead to unanticipated actions. The stage of the technological development, and the associated risks and implications, however, should be monitored by states and taken into consideration when testing new systems under AP I, art. 36.

References

- Earle, James. *Written Evidence on behalf of the Association of Military Court Advocates on 'Droning for Britain: Legal and Ethical Issues Arising from the Use of Remotely Piloted Air Systems'*. UK Parliament, Session 2013-14. Available at <http://www.publications.parliament.uk/pa/cm201314/cmselect/cmdfence/writev/772/rpa06.htm>, accessed 27 March 2015.
- Gomez, F. J. G. 'The Law of Air Warfare'. *International Review of the Red Cross* 323 (1998): 347–63. Available at <https://www.icrc.org/eng/resources/documents/misc/57jpcl.htm>, accessed 27 March 2015.
- Heyns, Chris. *Extrajudicial, Summary or Arbitrary Executions*. UN Doc A/68/382. New York and Geneva: UN, 2013. Available at http://www.un.org/en/ga/search/view_doc.asp?symbol=A/68/382, accessed 27 March 2015.
- International Committee of the Red Cross (ICRC). 'New Technologies and the Modern Battlespace: Humanitarian Perspectives'. Panel discussion, 25 March 2014. Videorecording of the discussion available at <https://www.icrc.org/eng/resources/documents/event/2014/03-06-research-and-debate-inaugural-event.htm> accessed 27 March 2015.
- International Court of Justice (ICJ). *Case Concerning the Gabčíkovo-Nagymaros Project*, ICJ Judgment, 25 September 1997. Available at <http://www.icj-cij.org/docket/files/92/7375.pdf>, accessed 27 March 2015.
- International Law Commission (ILC). *Draft Articles on Responsibility of States for Internationally Wrongful Acts, with commentaries*. New York: UN, 2001.
- Kalshoven, F. *Reflections on the Laws of War: Collected Essays*. Leiden: Brill, 2007.
- Kolb, R. *Advanced Introduction to International Humanitarian Law*. Abingdon, UK: Edward Elgar Publishing, 2015.
- Melzer, Nils. *Human Rights Implications of the Usage of Drones and Unmanned Robots in Warfare*. Brussels: European Parliament, 2013. Available at [http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/410220/EXPO-DROI_ET\(2013\)410220_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/410220/EXPO-DROI_ET(2013)410220_EN.pdf), accessed 27 March 2015.
- Multinational Capability Development Campaign (MCDC). *Policy Guidance: Autonomy in Systems*. Norfolk, VA: Supreme Allied Commander Transformation HQ, 2014.

- Nudd, Tim. 'Anne-Hathaway and Julie Taymor team up on Off-Broadway Play', *People*, 22 January 2015. Available at <http://www.people.com/article/anne-hathaway-julie-taymor-grounded-broadway>, accessed 27 March 2015.
- Sassoli, M. et al. (ed.). *Un Droit dans la Guerre?* Vol. I. Geneva: ICRC, 2012. Available at <https://www.icrc.org/fre/resources/documents/publication/p0739.htm>, accessed 27 March 2015.
- UK Ministry of Defence (MOD). *Written evidence from the Ministry of Defence, 'Remote Control: Remotely Piloted Air Systems'*. UK Parliament, Session 2013-14. Available at <http://www.publications.parliament.uk/pa/cm201314/cmselect/cmdfence/writev/772/rpa01.pdf>, accessed 27 March 2015.
- UN General Assembly. Report of the International Law Commission. A/56/10. New York: UN, 2001.

The Varied Law of Autonomous Weapon Systems

Rebecca Crootof

Abstract

What law governs autonomous weapon systems? Those who have addressed this subject tend to focus on the law of armed conflict and debate whether it is sufficiently flexible to regulate such weaponry. But while it will undoubtedly be one of the more significant sources of states' legal obligations, the international laws applicable to the development or use of autonomous weapon systems are hardly limited to those rules regarding the conduct of hostilities. Other legal regimes—including international human rights law, the law of the sea, space law, and the law of state responsibility—may also be relevant to how states may lawfully create or employ autonomous weapon systems, resulting in a complex web of international governance.

Introduction

What law governs autonomous weapon systems? Those who have addressed this subject tend to focus on the law of armed conflict (LOAC) and debate whether it is flexible enough to regulate such weaponry. But while it will undoubtedly be one of the more significant sources of states' legal obligations, the international laws applicable to the development or use of autonomous weapon systems are hardly limited to those rules regarding the conduct of hostilities. Instead, obligations and prescriptions from multiple legal regimes interact to form a complex and evolving web of international governance.

This chapter addresses this gap in the conversation by highlighting other applicable treaty provisions and customary international law. After briefly reviewing how these two sources of international legal obligations operate and the importance of the LOAC, it examines other legal regimes – including international human rights law, the law of the sea, and space law – that, at least

at this point in time, may provide the most relevant additional guidance on the lawful design and deployment of autonomous weapon systems. It bears noting at the outset, however, that this chapter provides only brief, initial reviews of some germane legal regimes. Without specific facts, it is impossible to analyse thoroughly how each of these regimes will regulate any given autonomous weapon system. Nearly every area of international law might provide pertinent guidance, as long as the associated legal rules are applicable either to the weapon system itself or to the situation in which the weapon system is used.⁽¹⁾ Accordingly, this chapter cannot provide a comprehensive list of all potentially relevant legal regimes. Rather, its primary aim is to draw attention to the sheer variety of types of international law that should be considered.

Autonomous weapon systems also raise legal issues that have yet to be addressed by any existing international law. Some of these – like the question of how international criminal law will assign responsibility for war crimes committed by an autonomous weapon system – are unique to weapons that can make independent (and thus sometimes unpredictable) determinations. Other unresolved legal questions are not specific to autonomous weapon systems, but the principled use of such weaponry will necessitate answering them. For example, the use of drones for extraterritorial targeted killing missions has reignited an ongoing debate regarding the interaction between the LOAC and human rights law. Both academics and practitioners are discussing, with renewed vigour, questions regarding the extraterritorial application of human rights obligations and the appropriate geographic and temporal limitations of armed conflicts. Should autonomous weapon systems be similarly employed, their use will raise many of the same issues.

Until precedents specifically addressing these and other issues are developed, the proper application of existing international legal regulations will remain ambiguous. Accordingly, this chapter also discusses the potential role of alternative, non-legal sources of guidance and governance – ranging from forms of transnational dialogue to codes of conduct to domestic law – that may inform the development of the varied law of autonomous weapon systems.

1. 'Indeed, virtually any branch of international law may become relevant...if the concerned State delegates the required performance of a legal obligation from human agents to military [autonomous weapon systems]': MCDC 2014a.

What is an Autonomous Weapon System?

For the purposes of this chapter, and in accordance with the other work in this collection, an autonomous system is defined as one that employs ‘autonomous functioning’, which describes ‘the ability of a system, platform, or software, to complete a task without human intervention, using behaviours resulting from the interaction of computer programming with the external environment.’⁽²⁾ Meanwhile, a ‘weapon system’ is ‘[a] combination of one or more weapons with all related equipment, materials, services, personnel, and means of delivery and deployment (if applicable) required for self-sufficiency.’⁽³⁾ Accordingly, an autonomous weapon system is one that employs autonomous functioning.⁽⁴⁾

This definition encompasses a wide assortment of weaponry, including ones currently in use, ones in development, and ones that exist (thus far) only in science fiction. It would cover Russia’s Platform-M, a multi-purpose weapon system designed ‘for gathering intelligence, for discovering and eliminating stationary and mobile targets, for firepower support, for patrolling, and for guarding important sites.’⁽⁵⁾ The definition would also apply to the Israeli Guardium unmanned ground vehicle, a boxy, car-like robot used to patrol the Israel/Gaza border. The Guardium can be operated remotely or employ autonomous functions, ‘both driving itself and responding to obstacles and events,’⁽⁶⁾ and can conduct surveillance and use non-lethal or lethal force.⁽⁷⁾ Nor are autonomous weapon systems limited to ground-based vehicles: this definition would include the airborne Israeli Harpy Loitering Weapon, which

2. MCDC 2014b.

3. *DOD Dictionary of Military Terms*, http://www.dtic.mil/doctrine/dod_dictionary/, accessed 14 August 2014.

4. Haider 2014 notes that NATO defines ‘autonomous unmanned aircraft’ as those ‘capable of understanding higher-level intent and direction, sensing its environment, and, based on a set of rules and limitations, choosing from alternatives and taking actions to bring about an optimal but potentially unpredictable state without human input’. See also DOD (2012), which defines ‘autonomous weapon systems’ as ones that, ‘once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation’. Crotoft (2015) defines an ‘autonomous weapon system’ as ‘a weapon system that, based on conclusions derived from gathered information and preprogrammed constraints, is capable of independently selecting and engaging targets’.

5. Tarantola 2014.

6. May 2012.

7. *Defense Update* 2014.

independently ‘detects, attacks and destroys enemy radar emitters’,⁽⁸⁾ and encapsulated torpedo mines, a type of sea mine now being used by both Russia and China that, when activated by a passing ship, releases a torpedo that then engages a target.⁽⁹⁾

The definition would exclude only weapon systems that employ automated functions. Although automated (or automatic) weapon systems require no human intervention, they ‘operate using a fixed set of inputs, rules, and outputs, whose behaviour is deterministic and largely predictable.’⁽¹⁰⁾ A tripwire-triggered landmine, for example, would be a weapon system with automated, rather than autonomous, functions.

Given the extent to which autonomous weapon systems are employed today, it is crucial to clarify the laws regulating their use. Before delving into specific legal regimes, however, it is worth briefly discussing the building blocks of most international legal obligations: treaties and customary international law.

The Sources of International Legal Obligations

In most domestic legal systems, an authoritative lawmaker creates the law for the governed community. The international legal order has no such recognised authority. Instead, it is comprised of multiple legal obligations with varying weight, the two primary sources of which are treaties and customary international law.

Treaties are written documents memorialising agreements between states.⁽¹¹⁾ They may be contract-like arrangements between a few states clarifying specific legal rights and duties, or more constitutive documents that provide general guiding principles and aspire to universal participation. Treaty law is grounded in the concept of state consent; without it, at least doctrinally, a state’s treaty obligations cannot be created, modified, or terminated. Because treaties create binding obligations on a state only after that state voluntarily expresses

8. IAI 2014.

9. Scharre 2014.

10. MCDC 2014b, 5. Haider 2014 also notes that NATO distinguishes ‘autonomous’ from ‘automated’ aircraft on the grounds that the latter’s ‘actions and outcome are scripted and predictable’.

11. UN 1969, art. 2.

its consent to be bound, treaty law has an ‘opt-in’ character; it cannot create obligations for state parties absent their explicit assent.⁽¹²⁾

In contrast, customary international law derives from ‘evidence of a general practice accepted as law’.⁽¹³⁾ A rule of customary international law is recognised as existing when states generally engage in specific actions (the ‘state practice’ requirement) and accept that those actions are obligatory or permitted (the *opinio juris sive necessitatis* element). While treaty law is binding only on state parties to that treaty, customary international laws are binding on all states, subject to a limited ‘opt-out’ exception.⁽¹⁴⁾ Additionally, and importantly for new weaponry, customary international law has no extensive temporal requirement.⁽¹⁵⁾ Thus, although practices or norms are often described as ‘evolving’ into customary international law, that evolution can occur quite swiftly.

Treaties and customary international law obligations may influence each others’ development in a number of complex ways. Treaties might codify customary international law, lending it the legitimacy of the written word; alternatively, certain treaty provisions may evolve to become generally binding customary international law. Where they regulate the same subject – say, the LOAC – treaty and customary international law may operate at different levels of generality, such that one serves a clarifying or gap-filling function for the other. Alternatively, one form of law may conflict with the other, in which case the more specific law (the *lex specialis*) will usually apply. If both laws are equally specific, the later-in-time law (the *lex posterior*) prevails.

At present, no treaties specifically address autonomous weapon systems. States are employing such weaponry today and are thereby creating precedential state practice, but these actions have not yet congealed into customary rules. Thus, at least for now, autonomous weapon systems are regulated only indirectly by other legal regimes, the LOAC chief among them.

12. Ibid., art. 34.

13. ICJ 1945, art. 38.

14. As a norm evolves into controlling custom, a state may avoid becoming bound by steadfastly contesting the norm’s existence or its status as customary international law. However, this bar for persistent objection is quite high, and once a norm is recognized as customary international law, states cannot later opt out.

15. ‘[T]he passage of only a short period of time is not necessarily, or of itself, a bar to the formation of a new rule of international law’. ICJ 1969.

LOAC

There is a plethora of treaties codifying the LOAC. These include (but certainly are not limited to) treaties discussing appropriate methods and means of warfare, clarifying state obligations to civilians and other victims of armed conflicts, describing rules specific to naval and air warfare, outlining regulations protecting cultural property, and creating individual criminal liability for war crimes. States may also conclude bilateral or limited multi-lateral treaties relating to any aspect of the LOAC. Finally, there are numerous customary rules that circumscribe possible lawful state action.

The LOAC impose certain obligations on states even during peacetime. Most relevantly, states are prohibited from designing weapons that do not meet certain standards. Article 36 of Additional Protocol I (AP I) to the Geneva Conventions provides the general rule regarding the legal review of new weapons:

In the study, development, acquisition or adoption of a new weapon, means or method of war, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.⁽¹⁶⁾

There are 173 state parties to AP I, and many non-party weapon-producing states have recognised an obligation to conduct article 36 reviews. Additionally, many argue that article 36 codifies a rule of customary international law: as all states are prohibited under customary international law from using weapons with certain characteristics (discussed below), every state must conduct article 36-like reviews to avoid fielding unlawful weapons.⁽¹⁷⁾

Under article 36, states must ensure that new weapons will comply with all applicable international law, especially the LOAC. There are two primary

16. UN 1977 (AP I), art. 36.

17. See, e.g., ICRC, 2006, 4: 'The faithful and responsible application of its international law obligations would require a State to ensure that the new weapons, means and methods of warfare it develops or acquires will not violate these obligations.'

restrictions on new weaponry.⁽¹⁸⁾ First, a weapon must not be indiscriminate by nature: i.e., it must be capable of being used in a manner that discriminates between lawful targets (combatants and military objectives) and unlawful targets (civilians, civilian objects, and incapacitated combatants). Usually this means that the weapon can be directed at a lawful target and that its effects are not uncontrollable. Biological weapons are paradigmatic examples of forbidden indiscriminate weapons, both because they do not distinguish between civilians and combatants and because their effects cannot be controlled. Second, new weapons must not cause unnecessary suffering or superfluous injury. Bullets generally are not prohibited by this provision, but bullets tipped with poison or specifically designed to cause untreatable wounds would be.

There is no reason why autonomous weapon systems, as a class, cannot be used discriminately and without causing superfluous injuries – although any given design might not comply with these requirements.⁽¹⁹⁾ Proposed autonomous weapon systems should be evaluated at all stages of their research and development, and certainly prior to their deployment, to ensure compliance with these humanitarian objectives and legal requirements.⁽²⁰⁾

Because a weapon system's autonomous capabilities in carrying out different tasks may affect when and where it can be lawfully used, it is crucial that individuals deploying and operating an autonomous weapon system have a thorough understanding of its capabilities and limitations. The US Department of Defense has recognised this: its directive on the use of autonomous weapon systems requires that operators are 'trained in system capabilities, doctrine, and [tactics, techniques, and procedures] in order to exercise appropriate levels of human judgment in the use of force and employ systems with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable [rules of engagement]'.⁽²¹⁾

18. AP I, arts. 35(2), 51(4). See also Thurnher (2013), which discusses the applicability of these requirements to autonomous weapon systems. These rules are also widely recognized as binding customary international law. See ICRC, *Rule 70*, ND; ICRC, *Rule 71*, ND.

19. Crootof 2015, 1894.

20. Knuckey (2014) notes the general governmental agreement regarding the customary nature of article 36 and its applicability to autonomous weapon systems.

21. DOD 2012, 11.

Once hostilities commence, the LOAC governs the use of autonomous weapon systems. Since there is currently no special law of autonomous weapon systems, laws regulating weapons use generally will be equally applicable to autonomous weapon systems and to other, non-autonomous weaponry. For example, a US manual discussing the international law of air and missile warfare does not differentiate between human-piloted, remote-piloted, and autonomously piloted aircraft.⁽²²⁾

Rules for weapons use include the customary rules regulating individual attacks – especially the requirements of distinction, proportionality, and feasible precautions in attack⁽²³⁾ – and other treaty obligations of the responsible state. Many proponents of a ban on autonomous weapon systems are concerned that they will never be able to comply with these customary principles, especially the distinction and proportionality requirements.⁽²⁴⁾ The distinction requirement obligates parties to a conflict to distinguish at all times between lawful and unlawful targets.⁽²⁵⁾ Such challenges in target distinction have become increasingly complex in modern conflicts between states and non-state actors, where the latter's strategy often depends on blending in with civilians. The proportionality requirement, meanwhile, prohibits any individual attack that 'may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.'⁽²⁶⁾ Commanders authorising a given attack are responsible for conducting a proportionality analysis, which necessarily requires a subjective weighing of incomparable and unquantifiable probabilities.⁽²⁷⁾ Commanders therefore

22. HPCR 2010.

23. Thurnher (2013) discusses the applicability of these requirements to autonomous weapon systems.

24. See, e.g., HRW and IHRC 2014a.

25. AP I, art. 48 codifies the customary rule.

26. Ibid., art. 51(5)(b) codifies the customary rule.

27. Witt (2014) notes the difficulty of assigning military objectives and human suffering objective values, particularly given the relevance of the context within which the engagement occurs.

enjoy a great deal of discretion in subsequent evaluations of their decisions under a 'reasonable commander' standard.⁽²⁸⁾

While autonomous weapon systems could be used in indiscriminate and non-proportional ways – as could any other weapon – they do not inherently violate these customary principles by virtue of having autonomous capabilities. For example, non-lethal autonomous weapon systems, like an autonomous electromagnetic jammer, would not raise grave concerns under either principle.

Furthermore, lethal autonomous weapon systems in use today have been used in discriminate and proportional ways. While it is widely acknowledged that autonomous weapon systems are generally incapable of distinguishing between lawful and unlawful targets, this does not imply that certain such weaponry cannot be lawfully employed in appropriate circumstances.⁽²⁹⁾ Indeed, autonomous weapons systems are being discriminately employed today in areas where there is little risk to civilians and other unlawful targets, such as in the Korean demilitarised zone.⁽³⁰⁾ Similarly, autonomous weapon systems can be (and are being) used in compliance with the proportionality principle, although when such use will satisfy this requirement will necessarily require a fact-dependent and context-specific evaluation.⁽³¹⁾ Individuals deploying autonomous weapon systems must therefore have an adequate understanding of both the battlespace environment and the weapon system's abilities and limitations in order to make informed proportionality assessments.

In short, autonomous weapon systems can be and have been lawfully used in armed conflict, though like other weaponry their use will be circumscribed by treaty and customary international law. Accordingly, states should develop policies and procedures to ensure compliance with LOAC requirements.

28. See ICTY (2003), which states: '[I]n determining whether an attack was proportionate it is necessary to examine whether a reasonably well-informed person in the circumstances of the actual perpetrator, making reasonable use of the information available to him or her, could have expected excessive civilian casualties to result from the attack.'

29. Schmitt 2013.

30. Crootof 2015, 1873-76.

31. Ibid., 1879-79.

Additional International Legal Regimes

The international laws relevant to the creation or use of autonomous weapon systems are hardly limited to those rules regarding the conduct of hostilities, even in situations of armed conflict. Other treaty or customary norms – like those governing state action with regard to human rights, on the sea, in outer space, or assigning state responsibility for private actors' actions – may also affect how such weapon systems should be designed and deployed. As these legal regimes did not anticipate the possibility of autonomous weapon systems, however, they are applicable only insofar as such weaponry is employed in or utilises the regulated field.

International Human Rights Law

There are well over 100 different human rights treaties and numerous customary international laws relating to human rights, including protections of general rights (such as the rights to life or housing), protections of specific rights (such as women's or children's rights), and both general and specific prohibitions (on genocide, slavery, torture, and assorted types of discrimination).⁽³²⁾

The sheer variety of international human rights law renders it difficult to make generalisations regarding its applicability. States will be bound to the treaties they ratify, which are generally presumed to apply within the state's territorial boundaries. There is some debate as to whether a state's treaty obligations are binding on all extraterritorial state action, but there is a growing consensus that human rights obligations do apply whenever and wherever a state exercises 'effective control' over a region or individual.⁽³³⁾ Meanwhile, human rights norms that have attained the status of customary international law will bind all states, but determining which norms have attained that status requires a rule-specific evaluation. Some human rights norms – such as the prohibitions on genocide and slavery – are widely recognised as customary international law; others – such as the right to housing or health care – are contested.

Notwithstanding its widespread application, scholars have only recently begun discussing the applicability of international human rights law to

32. Hathaway 2002.

33. Hathaway et al. 2012.

autonomous weapon systems, usually in the context of an armed conflict.⁽³⁴⁾ There are, however, three distinct situations in which human rights law may be relevant to the use of autonomous weapon systems: in domestic law enforcement, in anti-terrorism actions and other operations in which states use force that do not rise to the level of armed conflict, and in armed conflicts.⁽³⁵⁾ In domestic law enforcement, states will be bound by their domestic laws, ratified human rights treaties, and customary human rights law. Determining the prevailing legal rule in the latter two situations is more complicated, at least to the extent that the LOAC modifies or conflicts with human rights law.

Theoretical attempts to reconcile the two legal regimes tend to fit into one of three models: (1) the Displacement Model, in which the LOAC is assumed to completely displace human rights law, (2) the Complementarity Model, which presumes that both bodies of law apply and may be interpreted to accord with each other, and (3) the Conflict Resolution Model, which assumes that the two bodies of law may generally be interpreted in light of each other, but which also employs a 'decision rule' for determining which law applies when the requirements of the two legal regimes cannot be credibly reconciled.⁽³⁶⁾ States, international organisations, and international tribunals have adopted diverse versions of one of these three models, which have real-world impacts on certain fundamental human rights.⁽³⁷⁾

Although analyses of what law applies in a given situation will be highly fact specific, it is possible to advance some general guidelines for policy makers regarding the applicability of human rights law. First, those employing autonomous weapon systems should be aware of the variety of relevant human rights law, especially domestically ratified treaties. These treaties and customary international law will govern domestic law enforcement scenarios and – most likely – those extraterritorial law enforcement actions where the

34. Asaro (2012) argues that both international human rights law and international humanitarian law prohibit the delegation of decisions to use lethal force to an automated process. O'Connell (2014) observes, in the context of a discussion on autonomous weapon systems, that 'certain human rights principles apply even during an armed conflict'. Docherty (2014) discusses an emerging state consensus on the applicability of human rights law to autonomous weapon systems. Finally, HRW and IHRC (2014b) considers human rights concerns in both armed conflicts and domestic law enforcement situations.

35. Heyns 2014.

36. Hathaway et al.

37. Ibid., 1926–35.

state exercises effective control in the region or over an individual. If states anticipate using autonomous weapon systems outside their borders, they should pre-emptively clarify their legal rationale regarding the extraterritorial application of their human rights treaty obligations. States expecting to use autonomous weapon systems in situations where the LOAC may modify or conflict with their human rights obligations should determine their stance regarding the relationship between the two legal regimes and construct their policies regarding the use of autonomous weapon systems accordingly. Finally, while ensuring that the use of autonomous weapon systems is consistent with human rights law will pose difficult legal questions in some situations, certain fundamental human rights protections and prohibitions will apply in all circumstances. Autonomous weapon systems may not be used to further genocide, for example, or to torture or rape.⁽³⁸⁾

Law of the Sea

Due to concerns about their ability to distinguish between lawful military targets and unlawful civilian targets, automated and autonomous weapon systems have been primarily used in locations where there is a low probability of civilian interaction. One such region is the sea, which explains why defensive weapon systems were installed on warships decades before equivalent technology was used on land.⁽³⁹⁾ Accordingly, treaties and customary international law governing state action on the sea will apply to these and future sea-based autonomous weapon systems.

The 1982 UN Convention on the Law of the Sea (UNCLOS) includes a number of provisions, many of which are recognised as stating customary international law, that apply to ships with mounted autonomous weapon systems and possibly to independent seafaring autonomous weapon systems.⁽⁴⁰⁾ These include articles 192–96, which outline state obligations to protect and preserve both the marine environment generally and specific areas, such as the seabed and ocean floor, and article 301’s general prohibition on threats or uses of force.

38. Cf. Carpenter (2014).

39. According to FAS (2014), the US Phalanx Close In Weapon System (CIWS) has been a ‘a mainstay self defense system aboard nearly every class of ship since the late 70s.’ A land-based cousin, the US Centurion, was originally deployed in 2005. See Stoner (2009).

40. See UN 1982.

In addition to providing that the high seas ‘shall be reserved for peaceful purposes’ (art. 88), UNCLOS sets forth a number of prohibitions applicable to ships equipped with autonomous weapon systems that wish to exercise rights to innocent and transit passage. Ships may exercise the right of innocent passage in another state’s territorial sea, provided that its activities are not ‘prejudicial to the peace, good order or security of the coastal state’ (art. 19). Prohibited activities include:

- (a) any threat or use of force against the sovereignty, territorial integrity, or political independence of the coastal state, or in any other manner in violation of the principles of international law embodied in the UN Charter;...
- (c) any act aimed at collecting information to the prejudice of the defence or security of the coastal state;
- (d) any act of propaganda aimed at affecting the defence or security of the coastal state; ...and
- (k) any act aimed at interfering with any systems of communication or any other facilities or installations of the coastal state (art. 19).

Ships and aircraft exercising their right to transit passage must similarly ‘refrain from any threat or use of force against the sovereignty, territorial integrity or political independence of States bordering the strait, or in any other manner in violation of the principles of international law embodied in the Charter of the United Nations’ (art. 39). During transit passage, ‘foreign ships...may not carry out any research or survey activities without the prior authorization of the States bordering [the] straits [used for international navigation]’ (art. 40).

While automated and autonomous weapon systems have long been used on warships, future autonomous weapon systems may themselves be warships.⁽⁴¹⁾ Should they be granted warship status, such systems would gain certain rights and associated obligations. Warships have complete immunity from any state other than its flag state (art. 95), are entitled to seize pirates (art. 107), may exercise the right to hot pursuit (art. 111), and may exercise certain enforcement powers (art. 224). Nor are warships subject to the UNCLOS provisions requiring preservation of the marine environment (art. 236). However, a warship’s flag

41. Cf. *ibid.*, art. 29 defines ‘warship’.

state bears ‘international responsibility for any loss or damage to the coastal State resulting from the non-compliance...with the laws and regulations of the coastal State concerning passage through the territorial sea or with the provisions of [UNCLOS] or other rules of international law’ (art. 31).

Space Law

Like the sea, outer space is an environment in which autonomous weapon systems may be used with relatively little risk to civilians or civilian objects (unless, of course, the weapon systems select and engage Earth-based targets from space, or malfunction and crash). It might well be lawful to deploy an autonomous weapon system in space that, due to an inability to be directed solely against permissible targets, could not be lawfully used on Earth. Furthermore, because space is a hostile environment for human beings, there is an added inducement to minimise the need for human operators by increasing the weapon systems’ autonomous capabilities. Given these incentives, states are likely to field space-based autonomous weapon systems.

While the law of space is still incipient, the 1967 Outer Space Treaty, other space law treaties, and assorted UN General Assembly declarations provide guidance that will govern the design, use, and state liability for space-based autonomous weapon systems. But the ‘ceiling’ of space law regulation is sky high – aside from a few plain prohibitions, it allows for a wide range of potential extraterrestrial autonomous weapon systems.

The Outer Space Treaty has been ratified by 103 states and signed by 25 more,⁽⁴²⁾ but it may be binding on all states as a codification of pre-existing customary international law.⁽⁴³⁾ Most importantly, the treaty prohibits the use of space for certain destructive purposes. It provides that ‘States Parties to the Treaty undertake not to place in orbit around the Earth any objects carrying nuclear weapons or any other kinds of weapons of mass destruction, install such weapons on celestial bodies, or station such weapons in outer space in any other manner’ (art. IV).⁽⁴⁴⁾ This provision begs the question of what

42. The roughly top 10 countries with the greatest space presence – China, France, India, Iran, Israel, Japan, Russia, South Korea, the United Kingdom, and the United States – have all either ratified or signed the treaty.

43. See, e.g., Vereshchetin and Danilenko (1985), which notes that the claims of non-party equatorial states that they are not bound by the treaty’s text have been roundly rejected by the majority of states.

44. UN 1967.

constitutes a ‘weapon of mass destruction’: while commonly presumed to include nuclear, radiological, chemical, and biological weapons, it is less clear what else is encompassed under that heading. Large-scale explosives might be considered weapons of mass destruction, as might an autonomous weapon system armed with them.⁽⁴⁵⁾ Additionally, an autonomous weapon system that cannot be controlled might itself be considered a weapon of mass destruction, especially if it is able to engage Earth-based targets.⁽⁴⁶⁾

The Outer Space Treaty also states that ‘the moon and other celestial bodies shall be used by all States Parties to the Treaty exclusively for peaceful purposes’ (art. 4). Article 4 further forbids the ‘testing of any type of weapons’ on such bodies. Accordingly, state parties may not employ autonomous weapon systems on the moon or other celestial bodies for military purposes. But while the moon and celestial bodies may not be so used, the voids between them may⁽⁴⁷⁾ – and have been, ‘as evidenced by the existence of earth-orbit military reconnaissance satellites, remote-sensing satellites, military global-positioning systems, and space-based aspects of an antiballistic missile system.’⁽⁴⁸⁾

State parties retain jurisdiction and control over objects they have launched into space, but they also bear international responsibility for their activities in space and are internationally liable to other state parties for damage caused by their space-based objects.⁽⁴⁹⁾ This principle would necessarily also apply to damage caused by space-based autonomous weapon systems, regardless of whether the damage was intended or due to a weapons malfunction.

Other space law treaties elaborate on these initial provisions. The Agreement on the Rescue of Astronauts, the Return of Astronauts and Return of Objects Launched into Outer Space details state party obligations to return recovered objects and the launching state’s obligation to reimburse related expenses (art. 5),⁽⁵⁰⁾ the Convention on International Liability for Damage Caused

45. For example, Use of Weapons of Mass Destruction, 18 U.S.C. § 2332a is an American statute that lists explosive devices as weapons of mass destruction.

46. Of course, states are unlikely to field uncontrollable weapons for practical reasons. Dahm (2012) states that there is currently no military disadvantage in keeping humans involved in decisions regarding engagement, and therefore no demand to delegate that step.

47. Cheng 1997.

48. Shackelford 2009.

49. UN 1967, arts. 6–8.

50. UN 1968.

by Space Objects provides rules for assessing state liability for their actions in outer space,⁽⁵¹⁾ the Convention on Registration of Objects Launched into Outer Space requires state parties to keep a registry of their launched objects and furnish certain information to the UN Secretary-General (arts. 2,4),⁽⁵²⁾ and the Moon Treaty proclaims that the moon is the ‘common heritage of all mankind’ and reiterates that it and other celestial bodies should be used only for peaceful purposes (arts. 3, 11).⁽⁵³⁾

The UN General Assembly has also adopted a number of declarations of legal principles and resolutions regarding states’ activities in outer space.⁽⁵⁴⁾ Two declarations may have particular relevance for autonomous weapon systems. The Principles Relating to Remote Sensing of the Earth from Outer Space, which some have argued have customary international law status,⁽⁵⁵⁾ provide guidance on how to minimise controversy related to weapon (or other) systems that gather and process information from space.⁽⁵⁶⁾ The Principles Relevant to the Use of Nuclear Power Sources in Outer Space may also be applicable to autonomous weapon systems insofar as such systems are nuclear powered.⁽⁵⁷⁾

Other International Legal Regimes

The aforementioned international legal regimes are just some of the many that may regulate the design and deployment of autonomous weapon systems; this subsection discusses a few additional ones that will likely prove relevant in the near future. Again, these examples are not meant to be all inclusive; rather, they are intended to highlight the variety of international legal regimes that must be considered when evaluating the lawfulness of an autonomous weapon system.

Weapon systems that use the electromagnetic spectrum or international telecommunications networks may be governed by telecommunications law, which is regulated by the International Telecommunications Union. The 193

51. UN 1972.

52. UN 1976.

53. UN 1979.

54. UN 2002.

55. See Lyall and Larsen 2009.

56. UN 2002, 38–41.

57. Ibid., 42–8.

member states are required to legislate against ‘harmful interference’ that ‘endangers the functioning of a radionavigation service or of other safety services or seriously degrades, obstructs or repeatedly interrupts a radiocommunication service’ (art. 6).⁽⁵⁸⁾ Member states must also preserve the secrecy of international correspondence, unless such secrecy violates domestic laws or international conventions (art. 37). Although member states ‘retain their entire freedom with regard to military radio installations,’ they ‘must, so far as possible, observe... the measures to be taken to prevent harmful interference’ (art. 48(1)-(2)). If military installations are used ‘in the service of public correspondence or other services governed by the Administrative Regulations, they must, in general, comply with the regulatory provisions for the conduct of such services’ (art. 48(3)). Accordingly, while autonomous weapon systems that use telecommunications networks will likely be included within the scope of the exception for military installations, when they are used for non-military purposes they must be capable of complying with applicable regulatory provisions.

To the extent that they interfere with non-military aviation, autonomous weapon systems might also be regulated by the 1944 Chicago Convention on International Civil Aviation, which has 191 state parties. The Chicago Convention requires that all states show ‘due regard for the safety of navigation of civil aircraft,’ and, as amended, forbids using weapons that target civilian aircraft in flight.⁽⁵⁹⁾ During wars or emergencies, states may lawfully disregard these provisions, but only if they first ‘notif[y] the fact to the Council,’ a specialised UN agency that coordinates and regulates international air travel (art. 89).

Customary international law may also provide guidance regarding the development and use of autonomous weapon systems. For example, the law of state responsibility is regulated primarily by customary international law.⁽⁶⁰⁾ Under these customary rules, states will likely bear responsibility for any internationally wrongful actions committed by their autonomous weapon

58. ITU 1992.

59. UN 1944, art. 3(d). See also Protocol Relating to an Amendment to the Convention on International Civil Aviation, 10 May 1984, 23 I.L.M. 705, which ‘reaffirm[s] the principle of non-use of weapons against civil aircraft in flight’.

60. The International Law Commission has attempted to codify these general rules, but they have not yet been formalized in a treaty. See ILC 2001. Particular treaty regimes, such as the General Agreement on Tariffs and Trade, have established their own specific rules relating to state responsibility.

systems, and they will have an independent customary duty to compensate injured states for such actions.⁽⁶¹⁾ Should private actors employ autonomous weapon systems to commit internationally wrongful acts, this body of law will also be useful in clarifying when a state bears international responsibility for the consequences.

Likewise, customary international law provides most of the rules regarding the jurisdictional immunities for states and their property.⁽⁶²⁾ Although there are specific exceptions, states generally enjoy broad immunity from the jurisdiction of another state's courts. Specifically, national courts usually grant foreign states immunity in civil proceedings for human rights abuses committed abroad.⁽⁶³⁾ Should an autonomous weapon system be involved in such an abuse, at present the customary law would likely shield the launching state from related civil actions.

* * * * *

Clearly, the development and usage of autonomous weapon systems are regulated by far more than just the LOAC. International human rights law, the law of the sea, space law, and many other international legal regimes must be considered in evaluating whether and how a given autonomous weapon system may be lawfully used.

But while the applicable rules are extensive, they are also far from complete. Autonomous weapon systems raise a number of questions that international law has not yet answered. Some of these are specific to the unique nature of such weaponry: for example, a weapon systems' autonomy raises difficult questions regarding who is to be held responsible should it commit a war crime. Other issues – like the geographic scope of a battlefield, or the relationship between the LOAC and international human rights law – have long troubled legal scholars and practitioners, but are raised with new urgency by the existence of autonomous weapon systems.

Nor is it possible to predict how state use of autonomous weapon systems might modify these and other legal regimes. International law – even treaty law – is not set in stone; rather, it constantly responds to new ideological,

61. Ibid.

62. Foakes and Wilmshurst 2005. An attempted codification of this customary law – the UN Convention on Jurisdictional Immunities of States and their Property – is not yet in force.

63. Ibid., 8.

political, and technological developments. For example, the customary law of state responsibility currently suggests that states will be held directly responsible for the actions of their autonomous weapon systems. If machines eventually attain a sufficiently high level of autonomy, however, they may become more akin to private actors than machines – at which point their actions may not be directly attributable to states.

In the absence of directly applicable treaty and customary international law, and when indirect sources of international regulation are in flux, other sources of guidance and governance can provide some direction as to how states may (or should) act.

Alternative International Sources of Governance and Guidance

In addition to treaty and customary international law, there are less formal sources of international legal obligations, which are often lumped together under the heading ‘soft law’. While such sources are not legally binding on states, they carry more weight than purely political commitments⁽⁶⁴⁾ and thus may provide useful guidance in analysing novel legal questions.⁽⁶⁵⁾ These soft law sources take a variety of forms, including common understandings derived from international and transnational dialogue, non-binding resolutions and declarations, and even codes of conduct. Other, less weighty sources of guidance include civil society reports and policy briefs, industry practice, and academic works. Finally, states may also contribute to ongoing, norm-building dialogues by publicising their domestic laws and policies.

These various sources of governance and guidance will usually be perceived as legitimate and relevant to the extent that they reflect an informal consensus regarding appropriate state behaviour. While the resulting common understandings are not formally binding, they may be highly influential – and, over time, they might evolve into customary international law or provide foundational norms that are later codified in a treaty.

64. Guzman and Meyer 2010.

65. Koh (2012) discusses US practices with regard to ‘nonlegal understandings’, ‘layered cooperation’, and ‘diplomatic law talk’.

International and Transnational Dialogue

In May 2014, representatives from over 80 states and from UN agencies, civil society, and other international and transnational organisations convened for a ‘Meeting of Experts on Lethal Autonomous Weapon Systems’.⁽⁶⁶⁾ After attending sessions on technical issues, ethical and sociological aspects, legal concerns, and operational and military matters, delegates noted that the meeting had usefully contributed to forming common understandings.⁽⁶⁷⁾ In April 2015, the UN hosted a similar, second Experts Meeting to continue the conversation.

Such forums for discussion have myriad benefits. States and non-state delegates can ‘share information and “best practices”,...seek to harmonize policies and oversight mechanisms,...coordinate enforcement practice, and...anticipate, prevent and resolve inter-national disputes’.⁽⁶⁸⁾ Not only can the conversations create powerful norms of state behaviour, they can also clarify where existing law is unclear or where there is a lack of consensus, which increases the chances that these ambiguities will be addressed.⁽⁶⁹⁾

Resolutions and Declarations

In a recently issued resolution primarily focused on the use of armed drones for targeted killing, the European Parliament called on European Union member states and the European Council to ‘ban the development, production and use of fully autonomous weapons which enable strikes to be carried out without human intervention’.⁽⁷⁰⁾

While this specific resolution is unlikely to have much practical effect – not least because such weapon systems have already been integrated into the armed forces of some EU member states and have proven uniquely effective⁽⁷¹⁾ – similar non-binding resolutions and declarations can be highly influential. For example, many provisions in the UN General Assembly’s Universal Declaration of Human

66. Simon-Michel 2014.

67. Ibid.

68. Marchant et al. 2011.

69. See Crootof 2015, 1896–97.

70. EP 2014. The European Parliament is the only directly elected institution in the European Union (EU), and it exercises the EU’s legislative function along with the Council of the European Union and the European Commission.

71. Crootof 2015.

Rights, including the right to life and prohibitions on slavery and torture, are now recognised as having customary international law status. Others, while not having obtained such status, are nonetheless widely cited as aspirational goals. Future resolutions and declarations on autonomous weapon systems may similarly influence state action by clarifying common aims or principles.

Codes of Conduct

Codes of conduct are ‘non-binding and often somewhat general guidelines defining responsible, ethical behaviour and which are intended to promote a culture of responsibility’.⁽⁷²⁾ They may be issued by governmental and non-governmental agencies, industry groups, companies, professional societies, or conglomerations of such entities,⁽⁷³⁾ and they may aim to describe appropriate behaviour, professional ethics, or best practices.⁽⁷⁴⁾

As with other forms of dialogue, the process of developing codes of conduct itself contributes to the growth of germane norms. Codes of conduct are particularly useful, as they are a concrete text codifying shared understandings that can be applied in specific situations. In this way, codes of conduct are similar to treaties and non-binding resolutions and declarations – but unlike broadly worded multilateral documents drafted by state representatives, codes of conduct are usually drafted by technical experts to address specific issues.⁽⁷⁵⁾ Additionally, and especially when compared with treaties, codes of conduct may be easily revised to address new developments.⁽⁷⁶⁾ Given that we do not yet fully understand the benefits and risks posed by autonomous weapon systems,⁽⁷⁷⁾ developing flexible codes of conduct may be preferable to negotiating a treaty.⁽⁷⁸⁾

72. Marchant et al., 307.

73. Ibid.

74. Ibid., 309.

75. Cf. Ibid., 310.

76. Ibid., 308 notes that the Foresight Institute’s guidelines on autonomous replicating nanosystems have been updated six times.

77. Ibid., 283–84 discusses the risks inherent in complex technologies.

78. Ibid., 306.

Reports, Policy Briefs, Industry Practice, and Academic Contributions

Numerous non-governmental organisations have issued reports and policy briefs related to the development and use of autonomous weapon systems. Unfortunately, the majority advance arguments for banning autonomous weapon systems entirely, rather than providing considered guidance on how they may be lawfully developed and deployed.⁽⁷⁹⁾ But civil society's heightened interest in the subject is driving a productive academic and practical discussion, which will likely clarify norms governing autonomous weapon systems. Indeed, civil society may be credited with generating state interest in the subject, which in turn led to the aforementioned Experts Meetings.

Non-governmental organisations, relevant industries, and scholars will continue to consider the challenges posed by autonomous weapon systems. By issuing reports, recommending policies, engaging in or condemning certain practices, or publishing relevant papers, they will promote a continued conversation and help develop norms of appropriate state conduct regarding the development and use of autonomous weapon systems.

Domestic Law

Domestic law carries doubled relevance for the governance of autonomous weapon systems. First, a state's internal law and policies regulating the use of autonomous weapon systems will regulate its domestic actors and, by extension, its state practice. Second, by publicising such laws and policies, states may encourage public debate on their reasoning and conclusions, which may help generate general principles to which all states can subscribe. The US Department of Defense's publicised policy on how 'semi-autonomous' and 'autonomous' weapon systems are to be evaluated and employed,⁽⁸⁰⁾ for example, has received little critique and its definitions have been widely adopted by the scholarly community.⁽⁸¹⁾ The UK Ministry of Defence's advisory paper,⁽⁸²⁾ however, has been roundly criticised as setting the bar for autonomy in weapon systems so high as to effectively preclude the policy's application.⁽⁸³⁾ In the interest

79. See, e.g., HRW and IHRC 2012.

80. DOD 2012.

81. See, e.g., HRW and IHRC (2013), which describes the policy as a 'positive step', albeit an incomplete one, toward a complete ban on such weaponry.

82. MOD 2011.

83. See, e.g., Sharkey 2012.

of similarly contributing to the ongoing conversation and developing shared understandings, policy makers should endeavour to publicise policies related to autonomous weapon systems whenever feasible.

Domestic law will likely play a particularly influential role in determining which individuals are responsible when an autonomous weapon system commits a war crime, as states are responsible for investigating and prosecuting such abuses.⁽⁸⁴⁾ In certain circumstances, it will be the system's operator; in others, the programmer; in still others, the manufacturer. Distinguishing between different types of violations and associated liability regimes will be a regulatory headache, and at least initially, one best addressed domestically. After different national approaches are tested and refined, states may apply successful domestic strategies at the international level.

Conclusion

Given that they challenge certain fundamental assumptions upon which different international legal regimes are grounded, autonomous weapon systems raise a number of questions that the laws on the books cannot answer. Soft law and other informal sources of guidance and governance may help address some of these ambiguities, but until these and other legal questions are resolved, the international law governing autonomous weapon systems will remain incomplete.

Yet there is a far broader range of international law governing the creation and use of autonomous weapon systems than is usually acknowledged. This chapter highlighted a few likely relevant legal regimes, but nearly any area of international law might provide pertinent guidance, depending on the components of the weapon system itself or how and where it is used.

Given that international law already provides extensive guidance regarding the development and use of autonomous weapon systems, a state conducting an article 36 legal review or considering whether to deploy such weaponry cannot limit its assessment to what is permitted by the LOAC. Instead, it must determine – preferably with the assistance of a broad array of diverse specialists – when and how a given weapon can be lawfully used, given all of the states' varied and intersecting international legal obligations.

84. See, e.g., UN 1949, arts. 49, 51; AP I, arts. 85–8.

References

- Asaro, Peter. 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making'. *International Review of the Red Cross* 94 (2012):687–709.
- Carpenter, Charli. "Robot Soldiers Would Never Rape": Un-packing the Myth of the Humanitarian War-Bot. *The Duck of Minerva*, 14 May 2014. Available at <http://www.whiteoliphant.com/duckofminerva/2014/05/robot-soldiers-would-never-rape-un-packing-the-myth-of-the-humanitarian-war-bot.html>, accessed 28 March 2015.
- Cheng, Bin. 1997. *Studies in International Space Law*. Oxford: Clarendon, 1997.
- Crootof, Rebecca. 'The Killer Robots Are Here: Legal and Policy Implications'. *Cardozo Law Review* 36 (2015): 1837-1915.
- Dahm, Werner J.A. 'Killer Drones are Science Fiction'. *Wall Street Journal*, 15 February 2012.
- Defense Update*. 'Enguard! Introducing the Guardium UGV'. 2014. Available at <http://defense-update.com/products/g/guardium.htm>, accessed 16 August 2014.
- Docherty, Bonnie. 'Guest Post: Taking on 'Killer Robots''. *Just Security*. 23 May 2014. <http://justsecurity.org/10732/guest-post-killer-robots/>, accessed 28 March 2015.
- European Parliament (EP). *Resolution on the Use of Armed Drones*. Brussels: European Parliament, 2014. Available at <http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P7-TA-2014-0172>, accessed 28 March 2015.
- Federation of American Scientists (FAS). 'MK 15 Phalanx Close-In Weapons System'. Available at <http://www.fas.org/man/dod-101/sys/ship/weaps/mk-15.htm>, accessed 16 August 2014.
- Foakes, Joanne and Wilmshurst, Elizabeth. *State Immunity: The United Nations Convention and Its Effect*. London: Chatham House, 2005.
- Guzman, Andrew T., and Meyer, Timothy L. 'International Soft Law'. *Journal of Legal Analysis* 2 (2010):171–225.

- Haider, Andre. Unmanned Aircraft Systems Operations in Contested Environments (2014). Available at http://www.dglr.de/fileadmin/inhalte/dglr/fb/q3/veranstaltungen/2013_uav_autonomie/UAS_OPS_in_Contested_Environments.pdf, accessed 20 April 2015.
- Hathaway, Oona A. 'Do Human Rights Treaties Make a Difference?' *Yale Law Journal* 111 (2002):1935–2042.
- Hathaway, Oona A., Crotoft, Rebecca, Levitz, Philip, Nix, Haley, Perdue, William, Purvis, Chelsea, and Spiegel, Julia. 2012. 'Which Law Governs During Armed Conflict? The Relationship Between International Humanitarian Law and Human Rights Law'. *Minnesota Law Review* 96 (2012):1883–1944.
- Heyns, Christof. 2014. *Autonomous Weapons Systems and Human Rights Law*. Presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons 13–16 May 2014, Geneva, Switzerland. Available at [http://www.unog.ch/80256EDD006B8954/%28httpAssets%29/DDB079530E4FFDDBC1257CF3003FFE4D/\\$file/Heyns_LAWS_otherlegal_2014.pdf](http://www.unog.ch/80256EDD006B8954/%28httpAssets%29/DDB079530E4FFDDBC1257CF3003FFE4D/$file/Heyns_LAWS_otherlegal_2014.pdf), accessed 28 March 2015.
- Human Rights Watch and International Human Rights Clinic, Harvard Law School (HRW and IHRC). *Losing Humanity: The Case Against Killer Robots*, 2012.
- Human Rights Watch and International Human Rights Clinic, Harvard Law School (HRW and IHRC). *Review of the 2012 US Policy on Autonomy in Weapon Systems*, 2013.
- Human Rights Watch and International Human Rights Clinic, Harvard Law School (HRW and IHRC). *Advancing the Debate on Killer Robots: 12 Key Arguments for a Preemptive Ban on Fully Autonomous Weapons*, 2014a.
- Human Rights Watch and International Human Rights Clinic, Harvard Law School (HRW and IHRC). *Sharking the Foundations: The Human Rights Implications of Killer Robots*, 2014b.
- International Committee of the Red Cross (ICRC). *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*. Geneva: ICRC, 2006. Available at https://www.icrc.org/eng/assets/files/other/icrc_002_0902.pdf, accessed 28 March 2015.

- International Committee of the Red Cross (ICRC). 'Rule 70. Weapons of a Nature to Cause Superfluous Injury or Unnecessary Suffering'. Customary IHL Database. Available at http://www.icrc.org/customary-ihl/eng/docs/v1_rul_rule70, accessed 12 January 2015.
- International Committee of the Red Cross (ICRC). 'Rule 71. Weapons That Are By Nature Indiscriminate'. Customary IHL Database. Available at https://www.icrc.org/customary-ihl/eng/docs/v1_rul_rule71, accessed 12 January 2015.
- International Court of Justice (ICJ). *Statute*, T.S. No. 993. 26 June 1945.
- International Court of Justice (ICJ). *North Sea Continental Shelf (F.R.G. v. Den., F.R.G. v. Neth.)*, 1969 I.C.J. 4.
- International Criminal Tribunal for the Former Yugoslavia (ICTY). *Prosecutor v. Galić*, Case No. IT-98-29-T, 5 December 2003.
- International Law Commission (ILC). *Draft Articles on Responsibility of States for Internationally Wrongful Acts, with commentaries*. New York: UN, 2001.
- International Telecommunications Union (ITU). *Constitution of the International Telecommunications Union*. 22 December 1992. Available at <http://itu.int/net/about/basic-texts/index.aspx>, accessed 28 March 2015.
- Israel Aerospace Industries (IAI). 'Harpy Loitering Weapon'. Available at <http://www.iai.co.il/2013/16143-16153-en/IAI.aspx>, accessed 17 August 2014.
- Knuckey, Sarah. 'Governments Conclude First (Ever) Debate on Autonomous Weapons: What Happened and What's Next'. *Just Security*. 16 May 2014. Available at <http://justsecurity.org/2014/05/16/autonomous-weapons-in-tergovernmental-meeting/>, accessed 28 March 2015.
- Koh, Harold Hongju. 'Twenty-First Century International Lawmaking'. *Georgetown Law Journal Online* 101 (2012):1–19.
- Lyall, Francis, and Larsen, Paul B. (eds). *Space Law: A Treatise*. Farnham, England and Burlington, VT: Ashgate, 2009.
- Marchant, Gary E., Allenby, Braden, Arkin, Ronald, Barrett, Edward T., Borenstein, Jason, Gaudet, Lyn M., Kittrie, Orde, Lin, Patrick, Lucas, George R., O'Meara, Richard, and Silberman, Jared. 'International Governance of Autonomous Military Robots'. *Columbia Science and Technology Law Review* 12 (2011):272–315.

- May, Adam. 'Phantom on the Fence.' *Israeli Defense Forces Blog*. 9 October 2012. Available at <http://www.idf.il/1283-17082-EN/Dover.aspx>, accessed 28 March 2015.
- Multinational Capability Development Campaign (MCDC). *Autonomous Systems Focus Area (AxS FA): Consolidated Report*. Norfolk, VA: Supreme Allied Commander Transformation HQ, 2014a.
- Multinational Capability Development Campaign (MCDC). *Definitional Work for Autonomous Systems*. Norfolk, VA: Supreme Allied Commander Transformation HQ, 2014b.
- O'Connell, Mary Ellen. 'Banning Autonomous Killing: The Legal and Ethical Requirement that Humans Make Near-Time Lethal Decisions.' In *The American Way of Bombing: How Legal and Ethical Norms Change*, edited by Matthew Evangelista and Henry Shue, 224–36. Ithaca, NY and London: Cornell University Press, 2014.
- Program on Humanitarian Policy and Conflict Research at Harvard University (HPCR). *Commentary on the HPCR Manual on International Law Applicable to Air and Missile Warfare*, 2010.
- Scharre, Paul. 2014. 'Autonomy, "Killer Robots", and Human Control in the Use of Force—Part I.' *Just Security*. 9 July 2014. Available at <http://justsecurity.org/12708/autonomy-killer-robots-human-control-force-part/>, accessed 28 March 2015.
- Schmitt, Michael. 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics.' *Harvard National Security Journal Features* (2013):1–37.
- Shackelford, Scott J. 'From Nuclear War to Net War: Analogizing Cyber Attacks in International Law.' *Berkeley Journal of International Law* 27 (2009):192–251.
- Sharkey, Noel. 'Automating Warfare: Lessons Learned from the Drones.' *Journal of Law, Information and Science* 21 (2012):140–54.
- Simon-Michel, Jean-Hugues. *Report of the 2014 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)*, 2014. Available at <http://www.unog>.

ch/80256EDD006B8954/%28httpAssets%29/350D9ABED1AFA515C-1257CF30047A8C7/\$file/Report_AdvancedVersion_10June.pdf, accessed 28 March 2015.

Stoner, Robert H. 'R2D2 With Attitude: The Story of the Phalanx Close-In Weapons System (CIWS)'. *NavWeaps*. 30 October 30 2009. Available at http://www.navweaps.com/index_tech/tech-103.htm, accessed 28 March 2015.

Tarantola, Andrew. 'Russia's Military is Getting Killer Wall-E Robot Soldiers in 2018'. *Gizmodo*. 15 July 2014. Available at <http://gizmodo.com/russias-military-is-getting-killer-wall-e-robot-soldier-1604674629>, accessed 28 March 2015.

Thurnher, Jeffrey S. 'The Law That Applies to Autonomous Weapon Systems'. *ASIL Insights*. 18 January 2013. <http://www.asil.org/insights/volume/17/issue/4/law-applies-autonomous-weapon-systems>, accessed 28 March 2015.

UK Ministry of Defence (MOD). Joint Doctrine Note 2/11: The U.K. Approach to Unmanned Aircraft Systems, 2011.

United Nations (UN). *Convention on International Civil Aviation*, 15 U.N.T.S. 295. 7 December 1944.

United Nations (UN). Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 75 U.N.T.S. 31. 12 August 1949.

United Nations (UN). Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, 610 U.N.T.S. 205. 27 January 1967.

United Nations (UN). Agreement on the Rescue of Astronauts, the Return of Astronauts and Return of Objects Launched into Outer Space, 672 U.N.T.S. 119. 3 December 1968.

United Nations (UN). *Vienna Convention on the Law of Treaties*, 1155 U.N.T.S. 331. 23 May 1969.

United Nations (UN). Convention on International Liability for Damage Caused by Space Objects, 961 U.N.T.S. 187. 29 March 1972.

United Nations (UN). Convention on Registration of Objects Launched Into Outer Space, 1023 U.N.T.S. 15. 15 September 1976.

- United Nations (UN). Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 1125 U.N.T.S. 3. 8 June 1977.
- United Nations (UN). Agreement Governing the Activities of States on the Moon and Other Celestial Bodies, 1363 U.N.T.S. 53. 5 December 1979.
- United Nations (UN). *Convention on the Law of the Sea*, 1833 U.N.T.S. 3. 10 December 1982.
- United Nations (UN). United Nations Treaties and Principles on Outer Space: Text of Treaties and Principles Governing the Activities of States in the Exploration and Use of Outer Space. Geneva: UN, 2002. Available at <http://www.unoosa.org/pdf/publications/STSPACE11E.pdf>, accessed 28 March 2015.
- US Department of Defense (DOD). Directive 3000.09: Autonomy in Weapon Systems, 2012.
- Vereshcetin, Vladlen S., and Danilenko, Gennady M. 'Custom as a Source of International Law of Outer Space'. *Journal of Space Law* 13 (1985):22–35.
- Witt, John Fabian. 'Two Conceptions of Suffering in War'. In *Knowing the Suffering of Others: Legal Perspectives on Pain and Its Meanings*, edited by Austin Sarat, 129–57. Tuscaloosa: University of Alabama Press, 2014.

Civilian Developments in Autonomous Vehicle Technology and their Civilian and Military Policy Implications

James M. Anderson and John M. Matsumura

Abstract

Recent and near-future civilian developments and uses of autonomous vehicle technology will likely shape future military operations. Specifically, the automobile industry now appears close to substantial change, engendered by autonomous, or ‘self-driving’, vehicle technologies. This technology offers the possibility of significant benefits to social welfare – saving lives; reducing crashes, congestion, fuel consumption, and pollution; increasing mobility for the disabled; and ultimately improving land use. However, there are important disadvantages of these technologies that may be ameliorated by careful policy making. There are also important privacy, communications, regulatory, and liability issues associated with this technology. This chapter reviews civilian autonomous vehicle developments, discusses the obstacles to realising the benefits of autonomous vehicle technology in the civilian sector both now and in the future, and proposes possible policy solutions to these obstacles. It also examines how lessons learned in the civilian sector could improve how the military does business in the autonomous vehicles domain, with the prospect of enhancing the efficiency and/or effectiveness of future military applications.

Introduction

The General Motors *Futurama* exhibit presented at the 1939 World’s Fair in New York piqued the collective American and world imagination. Among other wonders, it promised that the United States would have an automated highway system and foretold the coming of a fundamental revolution in the surface transportation of passengers and freight. Today, nearly 75 years later,

the advances in autonomous vehicle (AV) technology (also known as automated driving systems) place us on the cusp of that revolution.

AVs have enormous potential to allow for the more productive use of time spent in a vehicle and to reduce crashes, costs of congestion, energy consumption, and pollution. They may also alter models of vehicle ownership and patterns of land use, and may create new markets and economic opportunities. Yet policy makers are only beginning to grapple with the immense changes that AVs portend. They face many policy questions, the answers to which will be influential in shaping the adoption and impact of AVs. These include everything from when (and whether) this technology should be permitted on the roads to the appropriate liability regime. In order to help policy makers maximise the net benefits of this technology, this chapter reports the results of a thorough literature review and interviews of approximately 30 stakeholders involved in the industry. It summarises recent civilian developments in autonomous vehicle technology, assesses their policy implications, and provides limited guidance to the military on the implications of the civilian developments of this technology. While this analysis is predominantly US-centric, the findings can easily be applied to other nations.

But what is autonomous vehicle technology, or road vehicle automation, as it is sometimes called? This technology is most easily conceptualised using a five-part continuum suggested by the National Highway Traffic Safety Administration (NHTSA), which describes different levels of automation:

- **Level 0:** The human driver is in complete control of all functions of the car.
- **Level 1:** One function (e.g., acceleration) is automated.
- **Level 2:** More than one function is automated at the same time (e.g., steering and acceleration), but the driver must remain constantly attentive.
- **Level 3:** The driving functions are sufficiently automated that the driver can safely engage in other activities.
- **Level 4:** The car can drive itself without a human driver.

While there is considerable excitement about the promise of road vehicle automation, it is important to distinguish between different levels of automation and the benefits they will provide. Many of the safety benefits of automation, for example, may occur at relatively low levels of automation, for example, by introducing automatic braking, while the mobility benefits are likely to be

realised only after the more daunting technical challenges of full automation are met.

Promise and Perils of Autonomous Vehicle Technology

AV technology has the potential to substantially affect safety, congestion, energy use, and, ultimately, land use. Conventional driving imposes not only costs borne by the driver (e.g., fuel, depreciation, insurance), but also substantial external costs, or ‘negative externalities’, on other people. For example, every additional driver increases congestion for all other drivers and increases the chance that another driver will have an accident. These externalities have been estimated at approximately 13 cents per mile.⁽¹⁾ If a hypothetical driver drives 10,000 miles, she imposes \$1,300 worth of costs on others, in addition to the costs she bears herself. AV technology has the potential to substantially reduce both the costs borne by the driver and these negative externalities, as discussed below.

Effect on Crashes

While the frequency of crashes has been gradually declining in the United States, such incidents remain a major public health problem. There were 32,179 fatalities in traffic crashes in 2013 in the United States,⁽²⁾ and approximately 2.3 million individuals were injured. Worldwide, the figures are much higher, with approximately 1.24 million traffic fatalities every year.⁽³⁾ The World Health Organization estimates that there are at least 20 non-fatal injuries for every fatality, leading to an overall annual rate of at least 24.8 million non-fatal injuries.

AV technology can dramatically reduce the frequency of crashes. The Insurance Institute for Highway Safety estimated that if all vehicles had forward collision and lane departure warning systems, sideview (blind spot) assist, and adaptive headlights, nearly a third of crashes and fatalities could be prevented.⁽⁴⁾ Automatic braking when the car detects an obstacle will also likely reduce a significant number of rear-end collisions.

1. Anderson et al. 2014.

2. NHTSA 2014a.

3. WHO 2013.

4. IIHS 2010.

Technologies that permit the car to be primarily responsible for driving (Level 4) will likely further reduce crash statistics, because driver error is responsible for a large proportion of crashes. This is particularly true given that 31% of the crash fatalities in the United States in 2013 involved alcohol use by one of the drivers.⁽⁵⁾ The overall social welfare benefits of vehicles that crash less frequently are significant, both for the United States and globally, and many of these benefits will go to those other than the purchasers of the vehicles with this technology.

Effect on Mobility

AV technology will also increase mobility for those who are currently unable or unwilling to drive. Level 4 AV technology, when the vehicle does not require a human driver, would enable transportation for the blind, disabled, or those too young to drive. The benefits for these groups would include independence, reduction in social isolation, and access to essential services.⁽⁶⁾ Some of these services are currently provided by mass transit or paratransit agencies, but each of these alternatives has significant disadvantages. Mass transit generally requires fixed routes that may not serve people where they live and work. Paratransit services are also expensive because they require a trained, salaried, human driver. Since these costs are generally borne by taxpayers, substituting less expensive AVs for paratransit services has the potential to improve social welfare.⁽⁷⁾

Effect on Traffic Congestion and its Costs

AV technology of Level 3 or higher is likely to substantially reduce the *cost* of congestion, since occupants of vehicles could undertake other activities. These reductions in the costs of congestion will benefit individual AV operators. On the one hand, reductions or increases in congestion itself are externalities that will affect all road users. A decreased cost of driving may lead to an increase in overall vehicle miles travelled (VMT), potentially increasing actual congestion.⁽⁸⁾

5. NHTSA 2014a.

6. Bradshaw-Martin and Easton 2014.

7. The growth and popularity of Uber and Lyft suggests the market for smartphone summoned mobility (Ryzik 2014).

8. Chase 2014.

On the other hand, the technology can also enable increased throughput on roads because of more-efficient vehicle operation and reduced delays from crashes.⁽⁹⁾ Thus, the overall effect of AV technology on congestion is uncertain.

Land Use

As noted above, AV technology of Level 3 or above will likely decrease the cost of time in a car because the driver will be able to engage in alternative activities. A related effect may be to increase commuter willingness to travel longer distances to and from work. This might cause people to live further from the urban core. Just as the rise of the automobile led to the emergence of suburbs and exurbs, so the introduction of AVs could lead to more dispersed and low-density patterns of land use surrounding metropolitan regions.

In metropolitan areas, however, it may lead to increased density as a result of the decreased need for proximate parking. One estimate concluded that approximately 31% of space in the central business districts of 41 major cities was devoted to parking.⁽¹⁰⁾ At Level 4, an AV could simply drop its passenger off and drive away to satellite parking areas. Another consideration is that AV-sharing programmes may decrease the rate of car ownership. In either event, fewer parking spaces would be necessary and would permit the greater development of cities.⁽¹¹⁾

AV technology may have different effects on land use in the developing world. Countries with limited existing vehicle infrastructure could ‘leapfrog’ to AV technology. Just as mobile phones allowed developing countries to skip the development of expensive land-line infrastructure, AV technology might permit countries to skip some aspects of conventional, human-driver-centred travel infrastructure.

Effect on Energy and Emissions

The overall effect of AV technology on energy use and pollution is uncertain, but it seems likely to decrease both. First, AV technology can improve fuel economy by 4–10% by accelerating and decelerating more smoothly than

9. Litman 2015.

10. Shoup 2005.

11. Lari, Douma, and Onyiah 2014

a human driver.⁽¹²⁾ Further improvements could be achieved by reducing the distance between vehicles and increasing roadway capacity. A platoon of closely spaced AVs that stops or slows down less often resembles a train, enabling lower peak speeds (improving fuel economy) but higher effective speeds (improving travel time). Over time, as the frequency of crashes is reduced, cars and trucks could be made much lighter, which would increase fuel economy even more.

AVs might reduce pollution by enabling the use of alternative fuels. If the decrease in the frequency of crashes allows lighter vehicles, many of the range issues that have limited the use of electric and other alternative vehicles would be diminished. A Level 4 vehicle could drop its owners off at a destination and then recharge or refuel on its own. One of the disadvantages of vehicles powered by electricity or fuel cells is the lack of a refuelling/recharging infrastructure, yet Level 4 AVs that can refuel themselves would permit a viable system with fewer refuelling stations.

Yet decreases in the cost of driving and additions to the pool of vehicle users (e.g., elderly, disabled, and those without driver's licenses) are likely to result in an increase in overall VMT. While it seems likely that the decline in fuel consumption and emissions would outweigh any such increase, it is not guaranteed.

Costs

While AV technology offers substantial potential benefits, there are also important costs, many of which, paradoxically, stem in part from its benefits. For example, since AV technology is likely to decrease the *cost* of congestion and increase fuel economy, it will also likely decrease the private cost of driving that a particular user incurs. Because of this decline (and because of the increase in mobility that AVs offer to the elderly or disabled), AV technology may increase total VMT, which in turn may lead to increases in the negative externalities of driving, including congestion and an increase in overall fuel consumption.

AV technologies may also disrupt existing institutions. By making proximate parking unnecessary, Level 4 AV technology may undermine the parking revenues that are an important and reliable source of funding for many cities.

12. NRC 2013.

Many smaller towns depend, in part, on traffic citations that are likely to become far less common. Alternative forms of revenue generation (e.g., user fees for city driving) might be able to replace these revenues, but not without some disruption. By providing a new level of mobility to some users, it may siphon riders (and support) away from public transit systems. Currently, one of the key attractions of public transit is riders' ability to undertake other tasks in transit. AV technology may erode this comparative advantage and lead to a decline in public transportation.⁽¹³⁾

Further, many jobs could be lost once drivers become unnecessary. Taxi, truck, and bus drivers may lose their livelihoods and professions. If crashes decline in frequency, an entire 'crash economy' of insurance companies, body shops, chiropractors, and others will be disrupted and many workers will be forced to find new jobs.

Overall, the benefits of AV technology – including decreased crashes, increased mobility, and increases in fuel economy – appear to outweigh the likely disadvantages and costs. However, further research would be useful to more precisely estimate these costs and benefits and whether they would accrue to the individual operator of the AV or to the public more generally. Such research would also help determine the optimal mixture of subsidies and taxes to help align the private and public costs and benefits of this technology.

State Regulatory Issues

A number of states, including Nevada, Florida, California, and Michigan (as well as Washington DC), have passed varying legislation regulating the use of AV technology. Measures have also been proposed in a number of other states, which may create conflicting regulatory requirements. It is also unclear whether such measures are necessary, given the absence of commercially available vehicles with this technology and the absence of reported problems to date with the use of this technology on public roads. Yet these proposals begin the conversations among the legislature, the public, and state regulatory agencies about this important and coming change in transportation.

13. Anderson et al. 2014.

Brief History and Current State of Civilian AVs

While futurists have envisioned vehicles that drive themselves for decades, research into AV technology can be divided into three phases. First, from approximately 1980 to 2003, university research centres worked on two visions of vehicle automation: (1) automated highways systems in which relatively ‘dumb’ vehicles relied on highway infrastructure to guide them and (2) AVs that did not require special roads.

Second, from 2003 to 2007, the US Defense Advanced Research Projects Agency (DARPA) held three ‘Grand Challenges’ that markedly accelerated advancements in AV technology. The first two were held in rural environments, while the third took place in an urban environment. Each of these spurred teams to develop the technology.

Third, private companies have more recently advanced AV technology. Google’s Driverless Car initiative has developed and tested a fleet of cars and initiated campaigns to demonstrate the applications of the technology – for example, through videos highlighting mobility offered to the blind.⁽¹⁴⁾ In 2014, Google introduced a driverless car that lacked even a steering wheel. In 2013, Audi and Toyota both unveiled their AV visions and research programmes. Nissan has also recently announced plans to sell an AV by 2020. Uber recently announced an initiative to research driverless car technology in cooperation with Carnegie Mellon University.⁽¹⁵⁾

Current State of Technology

Google’s vehicles, operating fully autonomously, have driven more than 800,000 miles without a crash attributable to the automation. Advanced sensors to gather information about the world, increasingly sophisticated algorithms to process sensor data and control the vehicle, and computational power to run them in real time has permitted this level of development.

In general, robotic systems, including AVs, use a ‘sense-plan-act’ design. In order to sense the environment, AVs use a combination of sensors, including light detection and ranging, radar, cameras, ultrasonic, and infrared. Sensors used in combination can complement one another and make up for any

14. Google 2012.

15. Trujillo 2015.

weaknesses in any single kind of sensor. While robotic systems are very good at collecting data about the environment, making sense of that data remains probably the hardest part of developing an ultra-reliable AV.

For localisation, the vehicles can use a combination of global positioning systems and inertial navigation systems (INS). Challenges remain here, as well, because these systems can be somewhat inaccurate in certain conditions. For example, error of up to a metre can occur in a 10-second period during which the system relies on INS. The Google cars have depended on the careful pre-mapping of areas in which the vehicles operate.⁽¹⁶⁾ At this point, it is not clear what suite of sensors is likely to emerge as the best combination of functionality and price – particularly for vehicles that function at Levels 3 and higher.

In order to permit autonomous operation without an alert back-up driver at the ready, the technology will need to degrade gracefully in order to avoid a catastrophe. For example, if some element of the system fails in the middle of a curve in busy traffic, there must be a sufficiently robust back-up system to ensure that the vehicle can manoeuvre to a safe stop. Developing this level of reliability is challenging.

The role of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in enabling AV operation also remains unclear. While this technology could ease the task of automated driving in many circumstances, it is not clear that it is necessary. Moreover, V2I might require substantial infrastructure investments – for example, if every traffic signal must be equipped with a radio for communicating with cars.

Partly as a result of all of these challenges, most (but not all) stakeholders anticipate that a ‘shared driving’ concept will be used on the first commercially available AVs, in which vehicles can drive autonomously in certain operating conditions – e.g., below a particular speed, only on certain kinds of roads, in certain driving conditions – and will revert to traditional, manual driving outside those boundaries or at the request of a human driver.

Human driver re-engagement will pose another key challenge. To experience the greatest benefits of the technology, human drivers will need to be able to engage in other tasks while the vehicle is driving autonomously. For safety, however, they will need to quickly re-engage (in a matter of seconds or

16. Madrigal 2014.

less) at the vehicle's request. Cognitive science research on distracted driving suggests this may be a significant safety challenge.⁽¹⁷⁾ Similarly, developing the appropriate mental models for human-machine collaboration may be a challenge for a technology that is widely available to the public.

In order to sidestep the difficult human-computer interaction issues, another model of AV development is based on smaller, low-speed (< 30 kph) vehicles that lack any means of steering or direct human control. In the spring of 2014, Google revealed such a prototype. The Citymobil project in Europe also uses such a model.⁽¹⁸⁾ This model of development anticipates full autonomy, but addresses safety concerns with low speed, instead of ever-present human backup.

Software upgrades also could pose challenges, as they might need to be backward compatible with earlier models of vehicles and sensor systems. Moreover, as more vehicle models offer autonomous driving features, software and other system upgrades will have to perform on increasingly diverse platforms, making reliability and quality assurance all the more challenging. System security is also a concern, to prevent viruses or malware from subverting the proper functioning of vehicles' systems.

State transportation departments may need to anticipate the use of vastly different kinds of AVs operating on roadways. This may pose challenges for the registration and requirements necessary for the vehicles to operate, and for the level of training that particular operators must have. One short-term action that might improve safety is requiring stricter conformance to road signage requirements, particularly those that involve construction or some alteration to the roadway. This would both aid human drivers and ease some of the perception requirements for AVs.

Role of Telematics and Communications

The transfer of data to and from moving vehicles is expected to play an important role in the development of AVs in several ways. First, vehicles may use cloud-based resources, for example by using continually updated 'maps' that rely in part on sensor data from other vehicles. Similarly, if one vehicle's sensors malfunction, it might be able to partly rely on another vehicle's sensors.

17. Lee 2014.

18. <http://www.citymobil2.eu/en/Downloads/Overview/>, accessed 28 March 2015.

Second, the federal government has supported the development of dedicated short-range communications (DSRC) applications that would allow V2V and V2I communications, and has reserved electromagnetic spectrum for this use. Third, nearly every stakeholder interviewed for this study noted the inevitable need for software updates that will require some form of communication. Finally, many stakeholders believe that increasingly sophisticated ‘infotainment’ content may occupy vehicle occupants when full-time driving is no longer necessary, and that this content may increase demand for AV technology.

A central ongoing policy issue is the future of DSRC. While licences became available in 2004, they have only been used in experimental and demonstration projects. Recently, the Federal Communications Commission announced in a Notice of Proposed Rulemaking that they were considering allowing unlicensed devices to share the spectrum allocated to DSRC for purposes unrelated to transportation use, e.g., expanding Wi-Fi. Numerous stakeholders interviewed for this study thought this might impede the development of AVs, despite the current lack of use of the spectrum allocated to DSRC. It now appears likely that NHTSA will eventually require DSRC to be included in new cars.⁽¹⁹⁾

Other communications policy issues include the need to update distracted driver laws and the need to harmonise developmental standards for communications platforms within automobiles, along with issues pertaining to data security, data ownership, and privacy.

Standards and Regulations

Government regulations and engineering standards are policy instruments used to address safety, health, environment, and other public concerns. *Regulations* are mandatory requirements developed by policy makers that are specified by law and are enforceable by the government. *Standards*, in contrast, are engineering criteria developed by the technology community that specify how a product should be designed or how it should perform. Both will play important roles in the emergence and development of AV technology.

The NHTSA is the primary federal regulator of safety, and typically enacts Federal Motor Vehicle Safety Standards that specify performance standards for a wide range of safety components, including specific crash test performance.

19. NHTSA 2014b.

NHTSA can also issue recalls and influence the marketplace through its New Car Assessment Program. However it has no jurisdiction over the operation of cars, actions of vehicle owners, maintenance, repair, or modifications that vehicle owners may make.

Voluntary standards are also likely to play an important role in standardising safety, assuring system compatibility, and easing some of the complex human–computer interaction problems by standardising the methods by which vehicles operate.

Liability Implications of Autonomous Vehicle Technology

The existing liability regime does not seem to present unusual concerns for owners or drivers of vehicles equipped with AV technologies. On the contrary, the decreased number of crashes and associated lower insurance costs that these technologies are expected to bring about will encourage drivers and automobile insurance companies to adopt these technologies.

In contrast, manufacturers' product liability may increase, which could lead to inefficient delays in adopting this technology. Manufacturers may be held responsible under several theories of liability. Warnings and consumer education will play a crucial role in managing manufacturer liability for these systems, but concerns may still slow the introduction of technologies likely to increase that liability, even if they are socially desirable.⁽²⁰⁾

One potential solution to this problem is to more fully integrate a cost-benefit analysis into the standard for liability in a way that considers the associated benefits. It is difficult to specify the appropriate sets of costs and benefits that should be considered, however, thus further research would be helpful.⁽²¹⁾

Manufacturers might be able to reduce these risks by changing the business model of vehicle manufacturing – e.g., offering the use of an automobile as a service rather than a product. Another approach would be for manufacturers to use technology to more closely monitor driver behaviour.⁽²²⁾

Policy makers could also take actions to reduce manufacturers' liability. Congress could explicitly pre-empt state tort law remedies, an approach that

20. Kalra 2009.

21. Anderson et al. 2014.

22. Smith 2014.

has some precedents. Congress could also create a reinsurance insurance backstop if manufacturers have trouble obtaining insurance for these risks. Finally, policy makers (including the courts) could adopt an irrefutable presumption of human control of a vehicle, to preserve the existing convention that a human driver is legally responsible for a vehicle. However, each of these approaches also has significant disadvantages, and it is unclear at this point whether any liability limitation is necessary.

Guidance for Civilian Policy Makers

A key overarching issue for civilian policy makers is the extent to which the positive externalities created by AV technologies will create a market failure. As detailed above, this technology has the potential to substantially benefit social welfare through its reduction of crashes and costs of congestion, decreased fuel consumption and emissions, increases in mobility, and, eventually, changes to land use. Some of these potential benefits will not accrue to the purchaser of the vehicle with this technology, but more generally to the public. Thus these positive externalities will not drive the economic demand for the technology. Similarly, additional VMT may cause negative externalities – including congestion – which may have a less than socially optimal outcome. A combination of subsidies and taxes might be useful to internalize these externalities, but more information is needed in this area.

Over-regulation also poses risks. Different states' attempts to regulate AV technology could result in a 'patchwork quilt' of incompatible requirements and regulations that would make it impossible to operate a vehicle with this technology in multiple states.

Historically, initial vehicle performance has been tested federally by NHTSA, while driver performance and annual inspections are monitored by state entities. Since an AV is the driver, but the human may be required to intervene in certain ways and under certain circumstances, this division of roles could become complicated.

Liability concerns may also slow the introduction of this technology. These might be addressed by a variety of policy maker approaches, including tort pre-emption, a federal insurance backstop, the incorporation of a long-term cost-benefit analysis in the legal standard for reasonableness, or an approach that continues to assign liability to the human operator of the vehicle.

Overall, the guiding principle for policy makers should be that AV technology ought to be permitted if and when it is superior to average human drivers. For example, safety regulations and liability rules should be designed with this overarching guiding principle in mind. Similarly, this principle can provide some guidance to judges struggling with whether a particular design decision was reasonable in the context of a product liability lawsuit.

AV technology has considerable promise for improving social welfare, but will require careful policy making at the state and federal level to maximise its promise. Policy maker intervention to align the private and public costs of this technology may be justified once its costs and benefits are better understood through further research and experience.

Legal theorist Oona Hathaway has argued that law is path dependent: the existing legal regime depends critically on decisions made earlier. She warns that the public may become 'locked in' to an unnecessarily inefficient system.⁽²³⁾ She notes that the QWERTY keyboard is a famous example of an arguably subpar technology that became locked in as a result of the path dependence of technological and economic development. In the field of AV technology, law and policy will play a critical role in shaping the paths of technological development and deployment. An early case, regulation, or other policy (or lack thereof) could permanently shape the development of this technology. These pathways may influence the course of development in this field for a long time, so it is important that policy makers get it as right as possible.

Unfortunately, this is quite hard. While AV technology appears to be a way of improving social welfare, we are still at a very early stage of development. As Yogi Berra is said to have noted, 'It is tough to make predictions, especially about the future.'⁽²⁴⁾ At this stage, there are many more questions than answers, and much work remains to be done. At some point, civilian policy maker intervention to align the private and public costs of this technology may be justified. But at this point, aggressive regulatory action is probably premature and would probably do more harm than good.

23. Hathaway 2001.

24. A similar sentiment was expressed in Steincke (1948), quoting it as a witticism used in the Danish Parliament in the late 1930s.

Brief History of Military Autonomous Vehicles

While there has been remarkable progress in AVs in the civilian sector in recent years, which presents areas of opportunity for leveraging in the military, it is not at all clear that civilian applications will precede the implementation of military AV applications, given some of the key differences between the two domains.⁽²⁵⁾ From a technological standpoint, many of the capabilities currently envisioned for civilian AVs were initially explored through military-funded research and development programmes. While the DARPA Grand Challenges mentioned above are often identified as the beginning point of the commercialisation of AVs, early military efforts go back over three decades, to the DARPA and Army Research Laboratory Demo Series of unmanned ground vehicle experiments, in which multi-mode sensors were linked to portable computers that ran algorithms to control actuators, which in turn governed the speed and direction of unmanned high-mobility multi-purpose wheeled vehicle platforms.

Over the years, this technology has evolved into many other customised autonomous platforms such as the Army Research Lab's experimental unmanned vehicle, the mobile detection assessment and response platform, which is being used on a small scale today. The Tank-Automotive Research, Development, and Engineering Center has created and rigorously tested many autonomous variants of 'optionally manned platforms'. Essentially, a wide range of military initiatives has been in existence for many years, and some of these capabilities have provided the foundation for commercial AVs. Thus, given its considerable head start in AVs, why has the military not yet fielded AVs on a large scale for mainstream use?

Operational Differences

Despite a common set of technologies, the challenges of implementing AVs in the military and civilian sectors can be considerably different. The RAND Corporation compared the complexity of commercial and military AV applications⁽²⁶⁾ and found that one of the key advantages for the commercial sector is that AV applications can have more inherent 'structure' and can be

25. Some areas of opportunity exist in the legal and ethical development areas, whereas much of the military's focus has been on science and technology.

26. Matsumura et al. 2014.

better controlled. That is, the environment in which civilian applications are implemented can be selected to match the competency of the technology. In areas where the risk is too high, additional structure can be added to the environment. For example, the technology for autonomous crop harvesters has been successfully demonstrated. While not 100% fool-proof, risk can be mitigated by adding more structure to the area of operation, i.e., the farm. For example, electronic kill switches can be included in the design directly within the environment to ensure that autonomous systems do not go astray.

Military operations and environments can be highly variable and highly unstructured by comparison. The military typically cannot select its environment – it must ‘go where the fight is’. It may have significantly less opportunity to impose structure on (or ensure control of) the operating environment than the civilian sector; it can be extremely fluid, marked by rapid change (sometimes by design), which can dramatically increase the complexity and limit the available structure that can be imposed.

Another area in which civilian and military applications have traditionally differed has been in the amount of shared space between AVs and humans. In early civilian applications, the presence of high-risk humans, e.g., bystanders, could largely be avoided, simply through careful selection of the application and various access controls. For example, autonomous floor cleaning machines (e.g., large-scale Roomba-like systems) have been successfully implemented in the past by choosing when the AVs operate (e.g., the middle of the night when humans are not present) and restricting access to the area (i.e., by locking the doors).

Many notional concepts for military AVs envision a high degree of interaction between soldiers and robots by design; these are referred to as manned-unmanned (MUM) operations. Concepts put forward by the US Army’s Training and Doctrine Command envision relatively seamless MUM operations that share the same geographic space. While in theory this shared interaction can entail great benefits (e.g., robots shielding humans during combat), there is risk as well. Since military operations can occur in rough terrain, in urban environments, with varying obscurants such as smoke, and various obstacles and debris, they can be very complex from a technological perspective. These kinds of highly dynamic operations have posed a challenge to experiment with, test, and evaluate, in part due to the risks to soldiers.

Another difference between civilian and military requirements is the pace of operations. Civilian AV applications may be able to control the pace or rate of completion of autonomous actions in order to mitigate risk. While clearly not always prudent, this kind of trade-off option generally exists in the civilian sector. For example, if accidents associated with AVs could be reduced by slowing their speed limit, this alternative could be evaluated in the context of a cost-benefit analysis. In contrast, the military often requires high-paced operations, for example to get inside an adversary's observe, orient, decide, act (OODA) loop. In many instances, cycle time matters, and if the rate of completion of an AV task is slowed, it can jeopardise the success of the operation. While AVs can help reduce the cycle time for an OODA loop, in many instances the options for slowing the rate of completion of some tasks may have a detrimental effect on mission outcomes.

Legal and Ethical Differences

There are many other legal and ethical distinctions between civilian and military AV requirements. Unlike civilian applications, military applications may involve the deliberate use of lethal force against an adversary. Involving partly or fully autonomous systems in this process is ethically controversial, and is currently undergoing extensive public debate, including deliberations within the United Nations. While it is clear that autonomous systems have the potential to claim human lives, to what extent should this be allowed and, perhaps critically, what would be the next available outcome?⁽²⁷⁾ Ultimately, care must also be taken to ensure that the autonomous systems are consistent with the Laws of Armed Conflict and other applicable rules of engagement, which generally require distinguishing between combatants and non-combatants, proportionality, military necessity, and avoiding unnecessary suffering. These concerns are generally absent in the civilian context.⁽²⁸⁾

Yet many of the liability concerns are less present in the military context. As noted above, fear of increased liability on the part of automakers may slow the effective development and use of these systems. This concern is absent in the military context because of the liability protections provided to defence

27. Matsumura 2014.

28. CLAMO 2000; Rawcliffe 2007; US Army 2006; Arkin 2007.

contractors. However, similar to both the civilian and military domain challenges and trade spaces, there is a need for detailed analyses that show the effect of autonomy and provide a clearer sense of a *greater good*.

Conclusions

As civilian AVs continue to expand their boundaries of use to include open roads and highways, some of their challenges are beginning to resemble those that military AVs have faced. Open roads and highways offer only partially structured environments, have a fair degree of AV-to-human interaction, and may have to operate at predefined speeds, unless policy and current laws are changed. These are tenants of military AVs, albeit at a more restrictive overall level. And the test and evaluation criteria for AVs in these environments may not be as diverse as those needed to certify an AV for military operations – in some ways the overall risk may be higher – given the ultimate scope and scale of the prospective civilian operation. Extremely low probability, but very high consequence failure modes must be fully understood before fielding civilian AVs on public roads. Of course, an alternative might be to field the capability in more contained private environments, where structure and control can be imposed.

While the civilian AVs opportunities are expanding, the military appears to be scaling back at least part of its vision for future AVs. One major AV initiative that is in the process of becoming a programme of record is the autonomous mobility applique system. This AV capability is now envisioned to be used as a *driver-assist capability* that may very well lead to driverless convoy vehicles at some point in the future. Initial applications of this capability will likely be in conservative settings such as long-haul convoy missions, where highways have been secured and access can be controlled.⁽²⁹⁾ In both the civilian and military sectors, the main technological challenge is to ensure that AV use in early implementation is fully understood. This will be a difficult task, given the magnitude of the overall changes involved. While operating accident free may be too high an expectation, fully understanding how things will change, and developing the framework to address these changes, is essential.

29. Matsumura et al. 2014.

The goal of this chapter was to provide an overview of the policy issues associated with the large-scale use of AVs faced by civilian policy makers, and to compare how these issues are (or could be) related to the military use of AVs, which have been in development for decades. In the past, these issues were less important simply because the technology was not ready for mainstream use. As the technology matures, the key policy issues associated with both the civilian and military applications of AVs will become increasingly important. More detailed policy guidance will be needed, which will likely have to evolve over time to accommodate the considerable changes that this capability entails. Given at least some convergence of the technology – and, more importantly, its application in the two sectors – there is an opportunity to leverage past policy development and lessons learned.

References

- Anderson, James, Kalra, Nidhi, Stanley, Karlyn D., Sorensen, Paul, Samaras, Constantine, and Oluwatola, Oluwatobi A. *Autonomous Vehicle Technology: A Guide for Policymakers*. Santa Monica, CA: RAND Corporation, 2014.
- Arkin, A.C. *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*. Georgia Institute of Technology, Technical Report GIT-GVU-07-11, 2007.
- Bradshaw-Martin, Heather, and Easton, Catherine. 'Autonomous or 'Driverless' Cars and Disability: A Legal and Ethical Analysis.' *Web Journal of Current Legal Issues* 20/3 (2014).
- Chase, Robin, 'Will a World of Driverless Cars Be Heaven or Hell?' *Citylab*, 3 April 2014. Available at <http://www.citylab.com/commute/2014/04/will-world-driverless-cars-be-heaven-or-hell/8784/>, accessed 11 February 2014.
- Center for Law and Military Operations (CLAMO). *Rules of Engagement (ROE) Handbook for Judge Advocates*, May 1, 2000.
- Google, 'Self-Driving Car Test: Steve Mahan'. *You Tube*, 28 March 2012. <http://www.youtube.com/watch?v=cdgQpa1pUUE>, accessed 28 March 2015.
- Hathaway, Oona, 'Path Dependence in the Law: The Course and Pattern of Legal Change in a Common Law System'. *Iowa Law Review* 86/2 (2001).
- Insurance Institute for Highway Safety (IIHS). 'New Estimates of Benefits of Crash Avoidance Features on Passenger Vehicles'. *Status Report* 45/5 (2010).
- Kalra, Nidhi, Anderson, James M., and Wachs, Martin. *Liability and Regulation of Autonomous Vehicle Technologies*. Berkeley: California PATH Research Report, 2009.
- Lari, Adeel, Douma, Frank, and Onyiah, Ify. *Self-Driving Vehicles: Current Status of Autonomous Vehicle Development and Minnesota Policy Implications*. University of Minnesota White Paper, August 2014.
- Lee, John D., 'Dynamics of Driver Distraction: The Process of Engaging and Disengaging'. *Annals of Advances in Automotive Medicine* 58 (2014):24–32.

- Litman, Todd, *Autonomous Vehicle Implementation Predictions; Implications for Transport Planning*. Victoria Transport Policy Institute, 2015. Available at <http://www.vtpi.org/avip.pdf>, accessed 28 March 2015.
- Madrigal, Alexis C., 'The Trick that Makes Google's Self-Driving Cars Work'. *The Atlantic*, 15 May 2014.
- Matsumura, John. 'War Robots Will Lessen Killing – Not Increase It', *Christian Science Monitor*, 18 October 2013. Available at <http://www.csmonitor.com/Commentary/Opinion/2013/1018/War-robots-will-lessen-killing-not-increase-it>, accessed 28 March 2015.
- Matsumura, John, Steeb, Randall, Lewis, Matthew W., Connor, Kathryn, Vail, Timothy, Boyer, Matthew E., DiAntonio, Steve, Osburg, Jan, Morganti, Kristy Gonzalez, McGee, James, and Steinberg, Paul S. *Assessing the Impact of Autonomous Robotic Systems on the Army's Force Structure*. Santa Monica, CA: RAND Corporation, 2014.
- National Highway Traffic Safety Administration (NHTSA). *Preliminary Statement of Policy Concerning Automated Vehicles*, 30 May 2013.
- National Highway Traffic Safety Administration (NHTSA). *Traffic Safety Facts – Research Note, 2013 Motor Vehicle Crashes: Overview*, 2014a.
- National Highway Traffic Safety Administration (NHTSA). 'Advance Notice of Proposed Rulemaking (ANPRM)'. *Federal Register* 79/161 (2014b).
- National Research Council (NRC). *Transitions to Alternative Vehicles and Fuels*, Washington, DC: National Academies Press, 2013.
- Rawcliffe, J. (ed.) The Judge Advocate General's Legal Center and School, International and Operational Law Department, *Operational Law Handbook*, 2007.
- Ryzik, Melena. 'How Uber is Changing Night Life in Los Angeles'. *The New York Times*, 31 October 2014.
- Shoup, Donald C., *The High Cost of Free Parking*, Chicago: Planner's Press, 2005.
- Smith, Bryant Walker, 'Proximity-Driven Liability'. *Georgetown Law Journal* 102 (2014).
- Steincke, K.K. *Farvel og Tak* (Goodbye and Thanks). Fremad, 1948.
- Trujillo, Mario. 'Uber to Start Work on Driverless Technology'. *The Hill*, 3 February 2015.

US Army. *Field Manual 6-22: Army Leadership*. 2006.

US Department of Transportation (DOT). 'National Transportation Statistics', 19 July 2013. Available at http://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/national_transportation_statistics/index.html, accessed 28 March 2015.

World Health Organization (WHO). *Global Status Report on Road Safety 2013 – Supporting a Decade of Action*. Geneva: WHO, 2013.

PART 3:

AUTONOMOUS SYSTEMS AND OPERATIONAL RISK

NATO's Targeting Process: Ensuring Human Control Over (and Lawful Use of) 'Autonomous' Weapons

Mark Roorda⁽¹⁾

Abstract

The prospect of the use of so-called autonomous weapon systems has raised significant legal and moral concerns. This chapter contributes to the debate by providing an alternative perspective to the current dominant focus on the technological capabilities of future weapons. The author argues that machines do not have to be able to distinguish and make proportionality calculations. No rule in IHL requires weapons to do so. It is 'merely' the effects of attack decisions that need to be in accordance with relevant norms. Human judgement is required to decide under what circumstances to allow a particular system – with its specific abilities – to operate. NATO's targeting process serves as an example how weapons may be used effectively and responsibly, partly by its incorporation of legal norms. The author concludes that weapons programmed to perform targeting tasks without direct human input may be lawfully used in many situations if the state employing the system would follow similar steps as described in NATO's targeting doctrine and if humans continue to make the critical decisions about when and how to employ the system given the conditions ruling at the time

1. The author wishes to thank Terry Gill, Paul Ducheine, Jeffrey Thurnher, and Merel Ekelhof for their valuable comments and suggestions.

Introduction

Autonomous weapon systems (AWS)⁽²⁾ have been defined as weapon systems that, once activated, can select and engage targets without further intervention by a human operator.⁽³⁾ A number of concerns have been raised about this envisioned type of weapon: (1) will they have the innate technological capability to perform the assessments that are required by the law of armed conflict (LOAC),⁽⁴⁾ (2) is it unethical to have a machine decide about the use of force, let alone life and death, and (3) who – or what – should be held accountable for wrongdoing that results from the weapon's autonomous functioning?⁽⁵⁾ These concerns have led to calls for a complete pre-emptive ban on AWS, and quite understandably so, if one considers the implications of machines using force without human direction, oversight, or control.⁽⁶⁾ However, the term 'AWS', its accepted definition, and the associated concerns commonly harbor a singular focus on the platform and its technological capabilities. This focus neglects the human responsibility to decide on a particular weapon's use. This is also reflected in descriptions of weapons as 'man out of the loop'. To consider a system to be autonomous, and to ascribe to it the power to *select* targets or *make*

2. This is the most commonly used term. Others include: lethal autonomous weapon system, used in the context of the Convention on Certain Conventional Weapons Expert Meeting; lethal autonomous robotics, used by the UN Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns; and killer robots, used by several non-governmental organisations that actively seek a pre-emptive ban.

3. For an overview of several definitions, see Geneva Academy of International Humanitarian Law and Human Rights 2014, 6.

4. It is generally acknowledged that LOAC is applicable to weapons with autonomous capabilities (see, for instance Melzer 2014, 3; Thurnher 2013; ICRC 2014, 8. LOAC rules on attack include provisions that: (1) distinction must be made between military targets and civilians and civilian objects; (2) attacks are prohibited that may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated; and (3) all feasible precautions must be taken to avoid, and in any event to minimise, incidental loss of civilian life, injury to civilians, and damage to civilian objects.

5. A fourth concern is that seemingly risk-free warfare would lower the general threshold for the international use of force, making war more likely. While this is a legitimate concern, nearly every (technological) development aimed at lowering the risk to own personnel – from bulletproof vests to armed drones – lowers, to a certain extent, the risks of warfare and thus the reluctance for military action. It is a political matter (of the UN Security Council or nations that consider the use of force in self-defence, on the invitation of another state, or on humanitarian grounds) to establish institutional safeguards to ensure that armed force is only used in accordance with international law, an issue that remains outside the scope of this chapter.

6. 'Both experts and layman have expressed a range of strong opinions about whether or not fully autonomous machines should be given the power to deliver lethal force without human supervision' (HRW and IHRC 2012, 35).

decisions without human interference, seems to presuppose a machine's self-governance. Such a perspective tends to lead to anthropomorphism, whereby the platform itself is expected to perform tasks that are required to be operated in accordance with operational and legal requirements. This perspective has shaped the debate, yet it might be counterproductive in assessing the efficacy and legitimacy of weapons that are programmed to perform targeting tasks without direct human input.⁽⁷⁾

This chapter argues that no weapon should be regarded as a single entity operating in a vacuum. Nor is adherence to relevant norms only realised during execution. Humans will determine what type of system to launch for what type of mission, and under what circumstances. It is this decision, and the planning considerations that inform it, that is essential to constrain the use of force and to ensure that operational and legal requirements are met.⁽⁸⁾

To exemplify how assessments during the planning stage may facilitate the effective and lawful use of force, the following section describes NATO's targeting process. A phase-by-phase analysis shows how operational and legal standards are incorporated into a process that culminates in one essential decision. The subsequent section will elaborate on this decision, which is conceptually central to the use of force by any means or method. Since it is a human decision, it directly triggers human responsibility. The implications of this analysis for the use of AWS will be discussed in the section that follows, after which conclusions are drawn.

NATO's Targeting Process

NATO's joint targeting doctrine details procedures aimed at effective and lawful engagement of targets.⁽⁹⁾ The process has developed over the course

7. The author proposes to refrain from using the term 'autonomous' in combination with 'system' and/or 'weapon'. On this point, see Chapter 2 by Andrew Williams on definitional issues.

8. Operational requirements may include: commander's goals, guidance, and intent; direct orders; tactical directives; rules of engagement; special instructions, etc. Legal requirements in armed conflict are primarily derived from the LOAC, but also from other applicable treaties, customs, and principles.

9. 'The targeting process is focused on achieving the JFC's objectives efficiently and effectively...as limited by applicable rules of engagement (ROE) and relevant international law, and strives to minimize collateral damage (CD) and fratricide' (NATO 2008, 1-4). AJP-3.9 (1-7) furthermore states that both those planning and authorising attacks and those carrying them out have a responsibility to apply international law.

of history,⁽¹⁰⁾ possibly even from before NATO was established, and has been formalised in recent decades. It has now become embedded in the training and execution of NATO's military operations. The doctrine defines joint targeting as: the process of determining the effects necessary to achieve the commander's goals,⁽¹¹⁾ identifying the actions necessary to create the desired effects based on the means available, selecting and prioritising targets, and synchronising fires with other military capabilities, and then assessing their cumulative effectiveness and taking remedial action if necessary.⁽¹²⁾

Most military operations start with political direction that, although not formally part of the targeting process, contains a number of restrictions for targeting purposes.⁽¹³⁾ NATO's principal political decision-making body, the North Atlantic Council (NAC), will issue strategic military goals and provide guidance to the Joint Force Commander (JFC) responsible for executing the campaign.⁽¹⁴⁾ The NAC will simultaneously pass down approved target sets,⁽¹⁵⁾ including possible priority targets,⁽¹⁶⁾ and the JFC will receive guidance on how these targets will be selected for attack.⁽¹⁷⁾ Although strategic-level goals are typically very general and the associated target sets are relatively broad, it is clear that targets selection is controlled from the top down. Moreover, the goals and the designation of target sets must not contravene legal requirements. When the NAC approves target sets that contain civilian installations, those objects and

10. See Osinga and Roorda forthcoming.

11. The term 'goal' is used in this chapter even though doctrine uses the term 'objective'. This is to prevent confusion with the LOAC usage of 'military objective' – in the sense of target – as described in Additional Protocol (AP) I, art. 52(2).

12. NATO 2008, 1-1. The term 'joint' points to the fact that it is a joint effort between all armed force components. US doctrine defines targeting as the process of selecting and prioritising targets and matching the appropriate response to them, considering operational requirements and capabilities (DOD 2013, I-1).

13. An operation's legal mandate can also influence the targeting options. See Gill forthcoming.

14. NATO 2008, 3-1. These objectives are passed down from the NAC through NATO's Military Committee and Strategic Command to the JFC.

15. NATO 2008, A-13. Target sets include, but are not limited to: ground forces and facilities, military leadership, military supply and storage, electric power, transportation/lines of communication, and industry.

16. The term is 'time-sensitive targets' (TST), which are targets that require immediate response because they pose (or will soon pose) a danger to friendly operations, or are highly lucrative, fleeting targets of opportunity (Ibid., 1-3). Possible TSTs include mobile rocket launchers, theatre ballistic missiles, naval vessels, and terrorist leadership.

17. Ibid., 3-1. Note that 'engagement' encompasses both violent and non-violent action, while 'attack' is defined as 'acts of violence against the adversary, whether in offence or in defence' (AP I, art. 49).

persons may only be attacked if they qualify as legitimate military objectives in accordance with the LOAC.⁽¹⁸⁾ The JFC then formulates operational-level goals that will serve as input for operational targeting. If at any time the JFC wishes to appoint targets that have not yet been approved by the NAC, such approval needs to be sought.

When the JFC has formulated his goals, the targeting process formally commences in a six-phase cycle: (1) analysis of the JFC's goals; (2) target development, validation, nomination, and prioritisation; (3) analysis of capabilities; (4) assignment of capabilities to be used; (5) planning and execution of the mission; and (6) assessment of the results.⁽¹⁹⁾ While doctrine describes these as distinct steps, in practice some phases are conducted simultaneously and the whole process is iterative, depending on specific (changing) circumstances.

Phase 1: Analysis of Commander's Goals

Those responsible for targeting will have to understand the JFC's goals, guidance, and intent in relation to the NAC-approved target sets, and will translate those inputs into the desired effects and concrete tasks that are logically related to the overall desired end state.⁽²⁰⁾ Within the targeting cycle, this is the first moment that concrete violent action and physical effects may be considered; a consideration that is informed by operational and legal requirements. In attack, a distinction should be made between military objectives and civilian objects and persons, and only military objectives may be the object of attack.⁽²¹⁾ This means that target sets can include both military forces and civilians, but that attacks resulting in effects such as 'destroy' or 'neutralise' can only be connected to the former, while engagements generating effects such as 'reinstate' or 'inform' can also be linked to the latter. Analysis furthermore encompasses

18. Ibid., A-13.

19. Ibid., 2-1.

20. Ibid., 2-2.

21. AP I, arts. 48, 51, and 52. See also Henckaerts and Doswald-Beck, rules 1 and 7. Military objectives are limited to objects that, by their nature, location, purpose, or use make an effective contribution to military action and whose total or partial destruction, capture, or neutralisation, in the circumstances ruling at the time, offers a definite military advantage. See AP I, art. 52(2) and Henckaerts and Doswald-Beck, rules 8–10. In so far as persons are concerned, objectives are limited to combatants, unless they have become *hors de combat*, and civilians who have lost their protected status by directly participating in hostilities. See Henckaerts and Doswald-Beck, rules 3–6 and 47.

formulating appropriate measures of performance and effectiveness – if the JFC has not already done so – which are prerequisites for measuring progress toward accomplishing those tasks, effects, and goals – a function that links phase 6 (assessment) back to phase 1.

Phase 2: Target Development, Validation, Nomination, and Prioritisation

Since the NAC's target sets are very broad, with the possible exception of priority targets, phase 2 is aimed at specifying those sets while satisfying applicable operational and legal norms. Target development involves the analysis by targeteers of the adversary's capabilities, the determination of the best targets to engage in order to achieve the designated goals, and the collection of essential information on the target to be able to engage it.⁽²²⁾ This is also the stage at which initial issues related to potential collateral damage and other undesired effects may become apparent, which will be considered during validation and nomination; these issues could lead to restrictions on the means and methods of engagement, to be considered in phase 3.

Validation encompasses a process to ensure compliance with the JFC's goals and guidance, to ensure compliance with international law and established rules of engagement, and to verify the accuracy and credibility of intelligence used to develop a target. If the desired effect on a target is a violent one, the 'law of attack'⁽²³⁾ requires that '[w]hen a choice is possible between several objectives for obtaining a similar military advantage, the objective to be selected shall be that the attack on which may be expected to cause the least danger to civilian lives and to civilian objects.'⁽²⁴⁾ After positive validation, targets are nominated via a process of obtaining approval through the proper channels, after which they are prioritised.⁽²⁵⁾

22. Since target development requires substantial amounts of intelligence analysing capacity, there are different levels of development, and only targets that are assigned the highest priority, depending on the objectives in a particular phase of the operation, are developed fully.

23. This is often referred to as the 'law of targeting'. See, for instance, Boothby 2012. However, the term 'law of attack' seems more appropriate, since targeting also encompasses, for instance, non-violent actions directed at neutrals (e.g., information operations), of which the regulating norms are presumably not derived from LOAC. On non-kinetic targeting, see Ducheine forthcoming.

24. AP I, art. 57(3).

25. NATO 2008, 2-3.

At any given moment, there are, both conceptually and practically, two categories of targets: *specific targets* that have been sufficiently developed and validated to be nominated for action against them (e.g., enemy vehicle X at location Y), and *targets belonging to a set* that as a whole has been selected for engagement, but within which specific targets have not yet been sufficiently developed and validated to be nominated for action against them (e.g., enemy vehicles of a certain type).

Targets that have been sufficiently developed are known to exist in an operational area, for which sufficiently detailed target data is available to either directly schedule their engagement or to be held on call to be prosecuted if the situation demands it. Engaging these targets is called deliberate targeting.⁽²⁶⁾ Those that have not yet been sufficiently developed are either known to exist in the operational area (for which additional data is needed to engage them), or are unknown to exist and emerge unexpectedly, but do meet criteria specific to operational goals. Whenever such targets are detected in the conduct of operations, additional data needs to be collected to nominate them prior to engagement, through a process called dynamic targeting.⁽²⁷⁾

26. Ibid., 1-2.

27. Ibid. Although, compared to deliberate targeting, dynamic targeting is more reactive and is often more time pressured, it remains offensive action that requires the approval of a target engagement authority. This should be distinguished from (tactical-level) combat engagement, in which units do not require approval for individual attacks in response to enemy engagement (which should, in turn, be distinguished from self-defence), although boundaries are not always easily identifiable.

Phase 3: Capability Analysis

In this phase, capabilities are analysed to assess what means and methods are available and which are the most appropriate to engage a prioritised target, depending on the desired effect. The availability and efficacy of weapons that are programmed to perform targeting tasks without direct human input is included in this assessment. The analysis will cover considering: whether a target's characteristics (e.g., location and type) require action of a particular component (land, maritime, air, or special operations forces); whether achieving the desired effects is best served by forceful or non-forceful means; and whether certain capabilities within those options are ill suited because of operational or legal requirements. When collateral damage concerns are involved, for instance, no mean or method may be used that is expected to cause excessive collateral damage relative to the anticipated military advantage,⁽²⁸⁾ and some means or methods might be discarded if others would better avoid or minimise collateral damage.⁽²⁹⁾

A mean or method will be rejected for particular targets if it either cannot achieve the desired effect, or if it is assessed that its use cannot meet other operational or legal standards. Conversely, if there are no apparent concerns at this stage and it can achieve the desired effect, the weapon will be included among the options for the commander to decide.

Phase 4: Capability Assignment

After having evaluated the outcome of the previous phases, the JFC, typically supported by a Joint Targeting Coordination Board, will match capabilities against targets and assign those capabilities accordingly. Doctrinally, a capability is matched to a selected target to achieve the desired effects; but in practice, due to a scarcity of means, it is not uncommon for targets to be selected based on the available weapons and their characteristics. Either way, it is a conscious decision to use a particular capability on a particular target. Any relevant constraints and restraints that have emerged from these four phases will be passed onto the assigned unit. Depending on the assigned unit's capacity to ensure

28. AP I, arts. 51(5)(b), 57(2)(iii).

29. The requirement to 'take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects' is enumerated in AP I, art. 57 (2)(a)(ii).

compliance with operational and legal standards during the execution phase, these limitations could range from very strict to very lenient. It is a matter of a commander's trust in his subordinates, but it remains his prerogative to impose stricter limitations if he has reason to believe that the assigned unit is unable to ensure compliance with the relevant obligations.

Phase 5: Mission Planning and Execution

The assigned unit will conduct mission planning in similar steps to phases 1 to 4, but on a more detailed, tactical level. Goals are re-evaluated, additional intelligence is collected, targets are further refined, and means and methods are chosen from within the assigned unit that are best suited to achieve the goals. Assessments may include: location, type, size, and material of target; civilian pattern of life; time of attack (day or night); weapon capabilities; weapon effects; directions of attack; munition fragmentation patterns; secondary explosions; infrastructural collateral concerns; personnel safety; and battlespace deconfliction measures. Again, in every judgement, operational and legal standards are taken into account, including the obligation to take feasible precautions in attack.⁽³⁰⁾ A commander will subsequently approve the operation, which includes approving the use of a particular mean and method against a specific target during execution. Moving into the execution stage, precautionary measures remain to be applied.

One moment is of crucial importance during this phase: the moment when a person performs (or does not perform) an act, after which it is irreversible that violent action will – or could – occur. In most cases this is the decision to fire or launch a weapon. For example, when a sniper pulls the trigger, this final decision – in a long series of decisions – after which humans can no longer influence its direct violent effects, is essential to ensure adherence to operational and legal requirements. This decision will be further examined in the next paragraph, after briefly mentioning the final phase in NATO's targeting cycle.

30. Precautionary measures include: doing everything feasible to ensure the target is a lawful military target, taking all feasible precautions in the choice of means and methods to avoid or minimise collateral damage, refraining from deciding to launch a disproportionate attack, cancelling an attack if the target is not a lawful military target, and giving effective advance warning if the circumstances permit. See AP I, art. 57. Soldiers are often issued ROE to ensure that they use force within the boundaries of LOAC.

Phase 6: Combat Assessment

After engagement, combat assessment is conducted to determine whether the desired effects have been achieved, using the measures of performance and effectiveness. These determinations feed back into phase 1 by adjusting goals and tasks. The assessment will inform a decision on re-engagement that will include a re-assessment of earlier considerations, as changes in circumstances dictate.

The Central Decision

Practically, in most cases the central decision mentioned in phase 5 is made on the verge of pulling a trigger or pressing a button. The person performing this act is the last one to verify that the resulting effects will be in accordance with operational and legal requirements. Conceptually, though, it is not so much the moment of pulling a trigger; it is the moment after which humans can no longer influence the direct violent effects. For some weapons, this moment occurs *after* a munition has been launched or fired. For example, when launching a cruise missile that is re-programmable in flight, the conceptual moment of ‘no return’ is the point at which re-programming is no longer feasible. For other weapons, this moment can occur *before* a munition is launched. Take the example of a weapon platform that is programmed to find and engage enemy tanks in a certain area. A munition will be launched when the system has positively identified an enemy tank – based on sensor input and a pre-programmed database – but the conceptual moment of ‘no return’ occurred at the launch of the platform. With the use of any means or method, there will be a moment after which control is lost over the direct outcome. It is this moment (and the human decision that allows the process to surpass this point) that should be scrutinised more closely.

The person making this judgement has to consider a crucial question: will the effects that are expected to result from this decision be in accordance with operational and legal requirements? The answer to this question is influenced by his knowledge of, at least, the following five factors: (1) the situation at the time of the decision; (2) the expected functioning and effects of the weapon (including the operator’s ability to aim it); (3) the possible changes in the situation between the decision and the effect taking place; (4) the accuracy of

the intelligence used for these assessments; and (5) the operational and legal requirements.⁽³¹⁾

The combination of these factors should lead to the conviction that the expected effects will remain within the boundaries of operational and legal standards. In this sense, a weapon does not have to be perfect so that it can be used in all circumstances; it merely needs to be appropriate for the situation in which it is intended to be used. This judgement will in some cases be relatively straightforward, while in others it might be tremendously difficult. Either way, if a decision maker is not convinced that the expected effects will remain within the boundaries of operational and legal standards, the decision should be postponed and planning should continue, or it should be cancelled. Depending on the circumstances, there is a vast range of planning options to decrease uncertainty and to facilitate this decision, including (combinations of): assigning more specific targets, limiting the type of targets to be engaged (preferably to those that have characteristics that do not resemble the characteristics of civilian objects or persons in the same area), limiting operations that lead to dynamic targeting situations, decreasing the size of the area of operations, limiting the timespan of the operation (especially the time between the decision and the effect), increasing intelligence collection, limiting the operation to areas in which intelligence on the situation is easily collected and assessed and the circumstances are less complex (i.e., uncluttered environments), imposing measures that provide clarity on the accuracy of intelligence, imposing measures that prevent changes to the current situation (e.g., establishing a perimeter), giving advance warning, testing assigned means, reprogramming assigned means, opting for other means, or any other feasible measures.

Implications for the Autonomous Weapon Debate

The non-governmental organisation ‘Article 36’ asserted correctly that: ‘[t]he exercise of control over the use of weapons, and, concomitant responsibility and accountability for consequences are fundamental to the governance of the use of force and to the protection of the human person.’⁽³²⁾ The previous two sections have shown that humans will exercise such control even when

31. In military terms, this knowledge is often referred to as ‘situational awareness’. For an interesting analysis of the difference between ‘awareness’ and ‘understanding’, see MOD (2010), 2-1 to 2-9.

32. Article 36 2013, 1.

using weapons that are programmed to perform targeting tasks without direct human input. The central decision after which human control is lost is preceded – when taking NATO procedure as example – by an extensive planning stage, during which humans will formulate overall goals, gather intelligence, select and develop targets, analyse the most suitable type of weapon, and decide under what circumstances and preconditions to employ a particular weapon. In this sense, weapon systems are never truly autonomous.⁽³³⁾ The argument that man will be out of the loop in targeting must be based on a very narrow definition of ‘the loop’. When Markus Wagner asserted that ‘AWS remove a combatant from the decision-making process over a particular situation altogether’,⁽³⁴⁾ he surely did not contemplate the wider decision-making process described above. Accepting this broader view would invalidate assertions that a system ‘selects’ the targets and should thus have the innate technological capabilities to distinguish between military targets and civilians, to make proportionality calculations, and to take precautionary measures in attack.⁽³⁵⁾ This view was already quite remarkable, since existing law does not require that compliance is guaranteed by a system itself – an expectation that would also seem unrealistic. LOAC ‘merely’ requires compliance, regardless whether a weapon’s capabilities

33. Unless machines would be programmed to take over the entire targeting process, from formulating goals to deciding when and where to launch. For speculations in this direction, see Roff 2014. This is a very disturbing, but fortunately also very unrealistic, idea. There would seem to be no compelling reason why states would develop weapons over which they have no control (aside from the question of whether it is technologically possible).

34. Wagner 2012, 115.

35. See, for instance, the following statements: The necessary processes for applying those rules ‘would seem extremely challenging to programme into an autonomous weapon system’, especially in light of contemporary dynamic conflict environments (ICRC 2014, 13). ‘[T]he underlying software must be able to determine whether a particular target is civilian or military in nature’, and ‘[w]ith respect to target selection the software would have to be designed so as to anticipate all potential decisions, either by programming them in or by designing decision rules that are capable of making such decisions with a myriad of factors to be weighed’ (Wagner 2012, 113 and 121). ‘[N]o autonomous robots...have the necessary sensing properties to allow for discrimination between combatants and innocents (see Sharkey 2009, 27). ‘[F]ully autonomous weapons would not have the ability to sense or interpret the difference between soldiers and civilians’ and they ‘would not possess human qualities necessary to assess an individual’s intentions, an assessment that is key to distinguishing targets’, and ‘the [proportionality] test requires more than a balancing of quantitative data, and a robot could not be programmed to duplicate the psychological processes in human judgment that are necessary to assess proportionality’ (HRW and IHRC 2012, 30–3).

ensure adherence or whether this is achieved by the manner in which the weapon is employed. In this respect, the rules are result oriented.⁽³⁶⁾

Moreover, when accepting the idea – perhaps the fact – that humans will remain in control by virtue of the central decision, the real question becomes: how well has the person deciding on the employment of a particular weapon assessed the implications of its use? This is also reflected in the now widely adopted concept that insists that human control must be ‘meaningful’.⁽³⁷⁾ Without judging whether ‘meaningful’ is the most appropriate term, it seems to require human control that results in *lawful* use of force – a result that (as has been shown) is particularly well served by applying a process similar to NATO’s targeting cycle.⁽³⁸⁾ Doctrine details:

While all reasonably feasible care must be taken at each stage of the targeting process, targeting decisions and actions are not legally judged based on perfection, or that of hindsight. Those involved need only take all those precautions that were reasonably feasible at the time of their decision or actions and in the circumstances prevailing at that time. However, this objective standard also means that recklessness, negligence and willful blindness provide no excuse to unlawful targeting’.⁽³⁹⁾

36. Thus, the debate on whether robots will ever be capable of fulfilling the distinction and proportionality assessments is both speculative and irrelevant.

37. For a brief overview of this concept, see UNIDIR 2014.

38. In the author’s view, the concept of meaningful human control should not serve to ban a particular technological development. Rather, it should inform standards for the human operator. This view also seems to be taken by Horowitz and Scharre 2015. Interestingly, their ‘essential components’ of meaningful human control (pp. 14–15) – with which the author agrees – are not unique to the use of so-called AWS, but are necessary elements in *lawfully* using any type of weapon. Hence, the concept of meaningful human control seems redundant to existing legal norms and possible operational procedures (of which NATO’s serve as an example) to implement those norms.

39. NATO 2008, 1-7. It is recommended that states using weapons programmed to perform targeting tasks without direct human input should issue guidance on these issues. See DOD 2012: ‘Persons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement (ROE).’ Whether, in the case of breaches of the LOAC, this also leads to legal responsibility of an individual (or state) is a matter of (international) criminal law (and the law of state responsibility) or other relevant national laws. Concerns about the possibility that no person will be held accountable for wrongdoing in the conduct of warfare are not unique to the use of weapons that are programmed to perform targeting tasks without direct human input. Such concerns should be focused on the substance of (international) criminal law, the state’s acceptance of international tribunal jurisdictions, and matters of transparency.

It may be argued that the current state of technological development would preclude the use of weapons programmed to perform certain targeting tasks ‘to such an extent as to render them ineffective for the large majority of operations.’⁽⁴⁰⁾ Conversely, complex systems for which it is difficult to assess how they would respond to certain (unforeseen) situations can also complicate judgement.⁽⁴¹⁾ In both cases, uncertainties can be mitigated by a vast range of planning safeguards.⁽⁴²⁾ Interestingly, this produces a paradox: the more sophisticated the weapon – for example one that can be used in complex missions within operational and legal norms – the more a commander could be inclined to restrict its use due to a lack of understanding of how the system would respond to circumstances (predictable or not) and what effects it would generate.

Conclusion

The planning stage is of immeasurable importance to shape military operations and to ensure effective and lawful conduct, which also applies in the use of weapons that are programmed to perform targeting tasks without direct human input. NATO’s process shows that many steps can be taken from the moment that violent action is conceived to be an option: target sets are approved, goals are formulated, targets are selected, weapon capabilities are analysed, and detailed tactical planning takes place. In each phase, operational and legal requirements are considered. Inevitably, there is a ‘point of no return’, after which humans can no longer influence the direct effects of the use of force. In the case of weapons that are programmed to perform targeting tasks without direct human input that might engage targets without further human involvement, this point is its launch. It is a human decision (and responsibility) to let the process develop past this point, and the person making this decision

40. Wagner 2012, 121–2.

41. An important implication is that military commanders and staff should be familiar, to a reasonable standard, with how a particular weapon would respond to potential situations. This could be easier with a weapon of low sophistication than for weapons programmed with perhaps millions of lines of code. It is unavoidable that there will be errors in programming lines. See McConnell (2004) for current standards and how to reduce the number of errors.

42. This means, for instance, that ‘[t]here may be situations in which an [AWS] could satisfy [the rule of distinction] with a considerably low level ability to distinguish between civilian and military targets’ (Thurnher 2013).

has to be satisfied, taking the associated risks into account, that the expected effects are in line with operational and legal requirements. Using a complex weapon will invariably lead to higher restrictions on, for instance, the size of the area of operation or the type of targets it may attack. Thus, while a weapon's capabilities are an important factor, it is the combination of multiple factors that should give the decision maker the necessary confidence: 'the crucial question does not seem to be whether new technologies are good or bad in themselves, but instead what are the circumstances of their use.'⁽⁴³⁾

Even with humans in control and ample possibilities to restrict the use of weapons to ensure adherence to operational and legal requirements, lawful conduct is not guaranteed. Any weapon can be used unlawfully. The key issue is that weapons programmed to perform targeting tasks without direct human input may be lawfully used in many situations if the state employing the system follows similar steps as described in NATO's targeting doctrine, and if humans continue to make the critical decisions about when and how to employ the system given the conditions ruling at the time.

43. ICRC 2011, 40.

References

- Article 36. *Structuring Debate on Autonomous Weapons Systems*, 2013. Available at <http://www.article36.org/wp-content/uploads/2013/06/Autonomous-weapons-memo-for-CCW.pdf>, accessed 28 March 2015.
- Boothby, William. *The Law of Targeting*. Oxford University Press, 2012.
- Ducheine, Paul. 'Non-Kinetic Capabilities: Complementing the Kinetic Prevalence to Targeting'. In P. Ducheine, M. Schmitt, and F. Osinga (eds), *Targeting: Challenges of Modern Warfare*. The Hague: Springer, forthcoming.
- Geneva Academy of International Humanitarian Law and Human Rights. *Autonomous Weapon Systems under International Law*, 2014.
- Gill, Terry. 'Some Considerations Concerning the Role of the *Ius ad Bellum* in Targeting'. In P. Ducheine, M. Schmitt, and F. Osinga (eds), *Targeting: Challenges of Modern Warfare*. The Hague: Springer, forthcoming.
- Henckaerts, Jean-Marie and Doswald-Beck, Louise. *Customary International Humanitarian Law*. Volume 1: Rules. Cambridge University Press, 2009.
- Horowitz, Michael, and Scharre, Paul. *Meaningful Human Control in Weapon Systems: A Primer*. Washington, DC: Center for a New American Security, 2015.
- Human Rights Watch and International Human Rights Clinic, Harvard Law School (HRW and IHRC). *Losing Humanity: The Case Against Killer Robots*, 2012.
- International Committee of the Red Cross (ICRC). *International Humanitarian Law and the Challenges of Contemporary Armed Conflicts/* Report prepared for the 31st International Conference of the Red Cross and Red Crescent, Geneva, October 2011.
- International Committee of the Red Cross (ICRC). Report of the ICRC Expert Meeting on Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects, 26–28 March 2014, Geneva.
- Osinga, Frans and Roorda, Mark. 'From Douhet to Drones, Air Warfare and the Evolution of Targeting' In P. Ducheine, M. Schmitt, and F. Osinga (eds), *Targeting: Challenges of Modern Warfare*. The Hague: Springer, forthcoming.
- McConnell, Steve. *Code Complete*. Redmond, WA: Microsoft Press, 2004.

- Melzer (ed.), Multinational Capability Development Campaign Legal Study, Consolidated Report of 6 August 2014.
- NATO. *Allied Joint Doctrine for Joint Targeting, AJP-3.9*. 2008.
- Roff, Heather. 'The Strategic Robot Problem: Lethal Autonomous Weapons in War'. *Journal of Military Ethics* 13/3 (2014): 211–27.
- Sharkey, Noel. Weapons of Indiscriminate Lethality, FifF-Kommunikation, January 2009: 26–9.
- Thurnher, Jeffrey S. 'The Law That Applies to Autonomous Weapon Systems'. *ASIL Insights*. 18 January 2013. <http://www.asil.org/insights/volume/17/issue/4/law-applies-autonomous-weapon-systems>, accessed 28 March 2015.
- UK Ministry of Defence (MOD). Joint Doctrine Publication 04, *Understanding*, 2010.
- US Department of Defense (DOD). Directive 3000.09, *Autonomy in Weapon Systems*, 2012.
- US Department of Defence (DOD). *Joint Publication for Joint Targeting, JP 3-60*. 2013.
- UN Institute for Disarmament Research (UNIDIR). *The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might Move the Discussion Forward*. Geneva: UNIDIR, 2014.
- Wagner, Markus. 'Autonomy in the Battlespace: Independently Operating Weapon Systems and the Law of Armed Conflict'. In Dan Saxon (ed.), *International Humanitarian Law and the Changing Technology of War*, Martinus Nijhoff Publishers, 2012.

Choosing the Level of Autonomy: Options and Constraints

Erik Theunissen and Brandon Suarez

Abstract

In the navigation domain, increased autonomy has been pursued for many years. Although an autonomous navigation system may seem the most desirable option from an operational perspective, there are many issues that may favour a concept of operation that still involves a human operator at the decision-making level. These include the difficulty of meeting and proving compliance with extremely high reliability requirements of complex, safety-critical autonomous functions and the associated development and certification challenges. An increase in system autonomy is typically achieved through an increased integration of automated functions in combination with a transfer of execution authority from the human operator to the system. As the trend in goal-based function integration and transfer of authority continues, such systems will approach autonomous operation. It is concluded that the best model for navigation will be one that combines autonomy and human decision making in order to leverage the advantages of each.

Introduction

From an observer's point of view, it is often not easy to distinguish between automated and autonomous systems. By using an automated flight control function, an unmanned aircraft can fly without requiring a pilot to provide stick and throttle inputs. Robots can play soccer without a human giving directions, let alone actually controlling the robot. In the case of unmanned aircraft, any deviation from a pre-planned trajectory or pre-planned manoeuvre will require approval from the pilot; by design, the automated flight control system does not have the authority to choose its own path. In contrast, the robots taking

part in the annual RoboCup challenge⁽¹⁾ have been designed with the capability and authority to act based on a set of goals, a pre-defined strategy, and the perception of the current situation – a high level of autonomy. Thus in order to prevent an incorrect classification, a discussion of system autonomy must go beyond the observer's perspective to include the design perspective. From the design perspective, required choices for the degree of automation at the function level, options for integrating automated functions, and the allocation of authority will provide a much better insight into how well the system is able to adapt to a changing, unpredictable environment to accomplish its goals.

In the navigation domain, increased function autonomy has been pursued for many years. Although an autonomous navigation system may seem the most desirable option from an operational perspective, there are many issues that may favour a concept of operation that still involves a human operator at the decision-making level. These include the difficulty of meeting and proving compliance with extremely high reliability requirements of complex, safety-critical autonomous functions and the associated development and certification challenges. Furthermore, 20 years ago, an aircraft accident in which an automated navigation system was used⁽²⁾ yielded questions concerning the liability of the operator versus the system manufacturer. An increased transfer of decision authority from the operator to the system will certainly raise questions regarding a (partial) transfer of liability from the operator to the manufacturer that must be addressed.

In the navigation domain, an increase in system autonomy is typically achieved through an increased integration of automated functions in combination with a transfer of execution authority from the human operator to the system. As the trend in goal-based function integration and transfer of authority continues, such systems will approach autonomous operation – the ability to adapt to a changing, unpredictable environment in order to accomplish a goal.

1. RoboCup is an annual international robotics competition founded in 1997 to promote robotics and artificial intelligence (AI) research by offering a publicly appealing, but formidable, challenge: to create a team of fully autonomous humanoid robot soccer players and in 2050 win a soccer game (complying with the official rules of the FIFA) against the winner of the most recent World Cup.

2. In the lawsuit following the 1995 accident of American Airlines flight 965, the aircraft operator went to court in an effort to force the companies that designed the flight management system and the navigation database to help pay monetary damages. In response, the companies claimed that these were not at fault, and that the pilots committed a series of errors.

This chapter starts with a discussion of an automated navigation function and references to research that has addressed related automation issues. Based on a distinction between (1) the different stages of information processing in a task and (2) the possible levels of automation, it describes how integration with other automated functions yields an increase in autonomy. Using examples, it illustrates how the use of the level of automation (LOA) classification allows a comparison of different (conceptual) options to reduce the dependency on the human operator for safety-critical functions. This will demonstrate that the best model for navigation will be one that combines autonomy and human decision making in order to leverage the advantages of each.

Automatic Navigation

Navigation is the science by which geometry, astronomy, radar, etc. are used to determine the position of a ship or aircraft and to direct its course. Navigation can be divided into two processes: estimating the current location and getting to the planned location. The latter process involves guidance and control. Guidance is the determination of a trajectory from a current position and velocity to a desired position and velocity that satisfies specified costs and constraints. Control is the determination and issuing of the commands to the vehicle actuators to implement the trajectory and preserve a stable feedback loop.

The navigation process can be described as a closed-loop control system. The inner loop is the one directly affected by the control actions, which result in accelerations and rotations of the vehicle. The inner loop is the one with the highest bandwidth, and typically requires continuous input for stabilisation. Closing this loop is also required to maintain/change the orientation to a particular reference needed to achieve the desired direction of travel. Systems to automatically close the stabilisation loop were designed and implemented as early as the 1970s.⁽³⁾ Later implementations also allowed the use of a reference provided by the directional (guidance) loop (flight path angle command systems). For navigation along a planned trajectory, the reference in the guidance loop follows from the difference between the planned and estimated positions. Computer-based closures of the stability, directional, and position loops enable a platform to travel from a specified origin to a specified destination along a

3. NASA 1974.

pre-defined route without any human input. A well-designed system is able to compensate for the effect of a range of external disturbances that impact acceleration and/or angular motions.

The capability to store a three-dimensional route, compute the required guidance to follow the route, and generate the associated control commands has existed for over 40 years. In the commercial aviation domain, such an automatic navigation capability was introduced with the Flight Management System (FMS), while weapon systems such as the Tomahawk cruise missile with a similar⁽⁴⁾ automatic navigation capability were introduced a decade earlier. The introduction of automation has also resulted in new types of failures. In commercial aircraft, a range of incidents and accidents has occurred in which the interaction of the pilot with the FMS and unwarranted reliance on the protection provided by the automation can be identified as the root cause. However, the general consensus is that automation has contributed significantly to the safety of aviation.

For over 30 years, a multitude of design philosophies has been suggested to solve the ‘interaction with automation’ problem. For example, in 1991 a human-centred automation philosophy was proposed,⁽⁵⁾ and in 1995 a crew-centred design philosophy was presented.⁽⁶⁾ In the context of the design of a system to support single-pilot operations, a form of human-centred automation coined ‘complementation’ (for ‘complementary automation’) was proposed.⁽⁷⁾ Many more have been suggested. Although these philosophies have certainly helped prevent bad designs, it has proven to be an illusion that human error can be eliminated by adhering to a particular design philosophy, or that systems can be designed that are sufficiently robust to correctly deal with human error. This is important to consider, because when designing systems with increased autonomy, a fair requirement would be to achieve an equivalent level of safety.

4. An important difference between the automatic navigation function as implemented in manned aircraft and in weapon systems such as the Tomahawk is the role of the human operator. Manned aircraft have a number of additional safety nets that must prevent mid-air collisions and controlled flight to the ground.

5. Billings 1991.

6. Palmer et al. 1995.

7. Schutte et al. 2007.

Similar to the increased use of automation in commercial aircraft, automation has been used to reduce the number of crew in military aircraft. Also, 'the combination of increasing military system and task complexity, in the face of inherent human limitations has set the stage for development of innovative system integration approaches involving the use of knowledge based technology'.⁽⁸⁾ In addition to pilot selectable automation, so-called adaptive aiding systems, such as those developed in the context of the Pilot's Associate (PA) programme,⁽⁹⁾ have been extensively explored. The PA concept relied on 'the use of Artificial Intelligence computers to create an Electronic Crewmember'.⁽¹⁰⁾ The PA is 'a decision support system that improves the pilot's situation awareness and augments his decision making capability'.⁽¹¹⁾ Statler concluded that 'this was an example of an advisory system that failed largely because insufficient attention was paid to determining what would be needed to make it acceptable to the user'.⁽¹²⁾ This was not the only failed development of an adaptive aiding system. Based on the results of a survey, Andes and Rouse concluded that 'unfortunately, despite the availability of a growing number of proof of concept studies, most aiding systems have been unsuccessful'.⁽¹³⁾

In this context, it is important to note that although much of the underlying automation is likely to be very similar, the purpose of the PA programme was to create a trusted associate, rather than a system capable of autonomous operation. The reported issues do not concern issues with the automation itself, but with human interaction with the automation. Especially in critical situations, pilots mistrusted the advice from automation if it differed from what they expected. What makes the design of interaction with automation particularly challenging is that there are many different ways in which an implementation can be designed to interact with a human operator. To better distinguish between options that differ in terms of human involvement and associated allocation of authority, Sheridan and Verplank introduced a classification scheme that

8. NATO 1991, iii.

9. DARPA and the US Air Force began the PA programme in 1986 to exploit AI technology in order to improve the mission effectiveness and survivability of advanced tactical aircraft (Smith and Casper 1991).

10. Emerson and Reising 1994.

11. Corrigan and Keller 1989.

12. Stattler 1993.

13. Andes and Rouse 1992.

distinguishes between 10 levels of automation (LOA), from ‘human performs all functions’ to ‘automation performs all functions’ (see Table 8.1 below.)⁽¹⁴⁾

Level of Automation

The navigation function comprises multiple, nested feedback loops. Especially for the outer loop, there are various ways to allocate the authority to use automation to close the loop. However, at the function level, the LOA classification may require a distinction to be made between sub-functions. Similar to the decomposition of the navigation function into position estimation, guidance computation, and the execution of control actions, other more generic schemes have been proposed to distinguish between the different phases of information processing that are involved in realising a certain goal. One example is the observe, orient, decide, act (OODA) loop introduced by Boyd.⁽¹⁵⁾ Another example is the four-stage model of human information processing (information acquisition, information analysis, decision selection, and action implementation).⁽¹⁶⁾ This model is used to illustrate how the LOA classification can be applied to more accurately compare automated functions, and how integration with a sufficient LOA can lead to an increase in autonomy. In research into unmanned systems, a comparable classification method has been proposed based on the OODA loop.⁽¹⁷⁾

14. Sheridan and Verplank 1978.

15. Boyd 1986.

16. Parasuraman, Sheridan, and Wickens 2000.

17. Clough 2002.

Table 8.1. The 10 Levels of Automation

10	The computer decides everything, acts autonomously, ignoring the human
9	informs the human only if it, the computer, decides to
8	informs the human only if asked, or
7	executes automatically, then necessarily informs the human, and
6	allows the human a restricted veto time before automatic execution, or
5	executes that suggestion if the human approves, or
4	suggests one alternative
3	narrows the selection down to a few, or
2	The computer offers a complete set of decision/action alternatives, or
1	The computer offers no assistance: human must take all decisions and actions

Source: Parasuraman, Sheridan, and Wickens 2000.

To illustrate how system autonomy increases by replacing the functions performed by a human operator, an automated navigation function is described as a set of control loops with each loop subdivided into the four stages of information processing. Figure 8.1 shows a simplified overview of the information flow between the various loops that together realise an automated navigation function with several options for human operator intervention. The LOA classification scheme has been applied to each stage of information processing.

The smiley face to the right of the automatic navigation system represents the human operator. The arrows, labelled UI 1 to UI 3, represent different types of user interfaces (UIs) that enable interaction with the system. UI 1 gives the navigation system the reference (planned) trajectory. The navigation system contains three separate processes, each of which is subdivided into the four information-processing stages. The number in each block indicates the LOA for the particular process. In the automatic navigation system illustrated, the upper Information Acquisition block obtains the planned trajectory, the actual position, and time and provides it to the Information Analysis block. Information Acquisition is performed on a periodic basis, without any human involvement (LOA 10). Each time new information is provided, the Information Analysis block tests whether the actual position matches the planned position

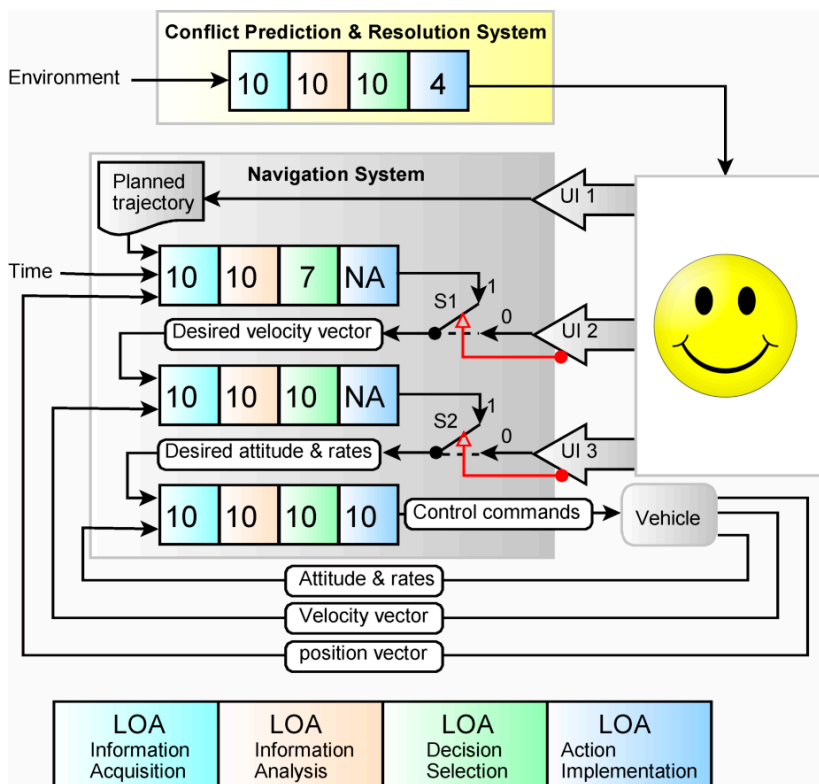


Figure 8.1. LOA Classification of Information Processing in an Automated Navigation Function

(LOA 10). Based on the difference, the Decision Selection process will compute a velocity vector that, if executed, will minimise the difference between the future planned and predicted positions. To allow a human operator to supervise the navigation process (LOA 7), the required information is made available by presenting relevant data on displays. In the upper block, there is no actual Action Implementation; the information is forwarded to the next process. Note that this information passes the switch S1, which is controlled by UI 2. This represents the human operator's opportunity to intervene in this control loop. Almost all commercial aircraft that are equipped with an automated navigation system provide the capability to intervene in the tracking of a pre-defined trajectory by specifying one or more components of the

velocity vector (e.g., by selecting a heading, speed, and/or rate of climb on the so-called mode control panel). In the process that inputs the actual and desired velocity vectors, all applicable stages of information processing have an LOA of 10. As in the previous block, there is no Action Implementation. The result is transferred via switch S2 to the bottom Information Acquisition block. Control commands are computed based on the difference between the desired and actual attitudes and rates. In this example an LOA of 10 is used in the Action Implementation block, but some systems provide (force) feedback (LOA 7) to the manual controls used by the pilot. Whereas switch S1 allows an operator to intervene by specifying elements of the desired velocity vector, switch S2 allows an operator to 'turn off' the automatic navigation and 'manually' control the vehicle. Some automation is less obvious/more hidden than other automation. Although perceived as 'manual control', in most modern aircraft the pilot still provides commands to an automated digital flight control system. This system computes the commands to the control surface actuators that will match the actual attitude and rate with the one commanded through inputs to the control stick/yoke. Envelope protection (to prevent unsafe aircraft states) is also typically provided in this loop.

Within a function that itself is automatic, autonomy can still exist at lower levels. For example, the guidance function may have sufficient authority to reconfigure itself (e.g., to adapt the configuration based on detected faults) without requiring the consent of the operator. Similarly, the sub-function that provides the upper Information Acquisition block with an estimate of the position of the aircraft is likely to have the autonomy to select the best sources for the required information. From the perspective of the user, similar to the earlier mentioned example of hidden automation, autonomy is not always obvious. Quite sophisticated adaptive systems were designed in the 1980s to 'manage a multisensory navigation suite, dynamically selecting the most appropriate navigation equipment to use in accordance with mission goals, mission phase, threat environment, equipment health, equipment availability, and battle damage'⁽¹⁸⁾ (in other words, an autonomous position estimation function). A mission planning sub-system with the capability to dynamically update the trajectory was designed in the context of the PA programme; it

18. Berning and Glasson 1990.

would perform route re-planning if required. If the Decision Selection and Action Execution LOAs were selected at 7 or higher, this would increase the level of autonomy of the navigation function.

In the architecture depicted in Figure 8.1, the pilot always has the authority to intervene in the automatic closure of the control, guidance and navigation loop. The Conflict Prediction & Resolution System block represents all systems that provide the pilot with information that would require him or her to intervene in the execution of the automated navigation function. An important characteristic of the architecture illustrated in Figure 8.1 is that the Action Implementation block of the Conflict Prediction & Resolution System has LOA of 4. Human operator action is required to input any resolution information that is provided to intervene in the navigation function. The system cannot perform this autonomously.

Safety

Clearly, the failure of an automatic navigation system can lead to unsafe situations. Regulations such as 14 CFR 25.1309 dictate that certain combinations of likelihood and consequences of failure are unacceptable. To meet the resulting requirements for automatic landing, fail-operational system architectures were introduced.⁽¹⁹⁾ Likewise, redundancy and dissimilarity are used to prevent single-point and common-mode failures. If a fault does occur, continued function availability is ensured through automatic fault detection, identification, and isolation. Furthermore, because safety can only be realised under specific operational conditions, requirements on the environment are imposed. The instrument landing system (ILS) has provided aircraft with the capability to automatically fly the final approach to the runway and land for 40 years.

However, even where no failure occurs, a system that is designed to automatically track a predefined trajectory is not automatically safe. With the ILS, a high integrity reference trajectory is established by means of a specifically designed antenna-beam pattern emanating from antennas on the ground close to the runway. A receiver on board the aircraft measures the deviation from the reference antenna beams and provides this information to the guidance system.

19. Charnley 1989.

This process eliminates the potential cause for an error in the definition of the reference trajectory. Safety is also based on the requirement that the runway is clear of obstacles and that no other aircraft comes close to the trajectory along which the aircraft is guided.

For automated navigation, an important assumption is that the potential hazards in the environment are sufficiently 'known' (i.e., that uncertainty can be kept below an acceptable maximum) and have been taken into account in specifying the trajectory. However, as a result of an error, the trajectory itself may be unsafe. On 20 December 1995, while en route to Cali (Colombia), a wrong waypoint name was entered into the FMS of a Boeing 757, and as a result the system generated an erroneous trajectory that was used as an input to the automatic navigation function. After the crew detected that the aircraft had turned in the wrong direction, they started a manoeuvre to return to the original path. However, this corrective manoeuvre caused the aircraft to crash into a mountain. The airline later went to court in an attempt to have part of the liability assigned to the manufacturer of the FMS and the company that developed the database software.

An error in an input to the directional control loop can cause the same type of accident. On 20 January 1992, the crew of an Airbus A320 assumed they had specified an angle of descent of 3.2 degrees, whereas in fact they had commanded a vertical velocity of 3,200 ft/minute (both would have to be specified at the level of UI 2 in Figure 8.1). The result was a much steeper descent that ended prematurely in the slope of a hill. Before this accident occurred, for the same aircraft type at least six incidents had been reported of situations in which pilots had entered a vertical speed when they had intended to specify an angle of descent. The user interface was later modified to reduce the likelihood of an error in the specification of the flight path angle.

To account for the possibility that both air traffic control and the pilot have failed to detect a collision hazard in a timely manner, last-minute safety nets are mandated. The Airborne Collision Avoidance System (ACAS) predicts collision hazard situations with other traffic and provides the pilot with guidance commands to increase the minimum separation. The Enhanced Terrain Awareness Warning System and the Ground Collision Avoidance System (GCAS) predict collision hazard situations with terrain and provide the pilot

with guidance commands needed to regain adequate separation. With few exceptions, these systems require the pilot to execute the advice generated by the system. This role of the human operator is the same as in the example in Figure 8.1: only the pilot has the authority to execute the advice generated by the conflict prediction and resolution function.

Authority of a System To Intervene

At present, only a few types of aircraft systems designed to detect and deal with unexpected situations have sufficient authority to directly intervene in the guidance loop. This integration enhances the automated navigation function with the capability to autonomously adapt to unexpected hazards such as traffic and/or terrain. One such system has resulted in a tighter integration of an existing automated navigation function with an existing airborne collision avoidance function. If the vertical navigation mode of the automatic navigation system is engaged (i.e., if the pilot has already delegated the authority to the system to automatically track the reference trajectory), the collision avoidance function has sufficient authority to command the flight control system to automatically execute the Resolution Advisory without requiring pilot action. The LOA of the Action Implementation has increased from 4 to 7.

Another example of a system that has the authority to intervene is the automatic ground-collision avoidance system (Auto-GCAS) that has been implemented on US F-16 aircraft.⁽²⁰⁾ This system has sufficient authority to input commands into the flight control system, even when the automatic navigation function is not engaged. It will over-ride the pilot's manual control inputs (or lack thereof). The first officially reported operational 'save' in which the system intervened during a mission took place in November 2014. In February 2015, more than 440 F-16s were already equipped with Auto-GCAS.

If systems such as ACAS and GCAS are integrated with the guidance or control loop and have sufficient authority to change the references used in these loops, from the perspective of the navigation function the dependency on the pilot has reduced. This represents an increase in autonomy. Figure 8.2 uses the overview presented in Figure 8.1 to illustrate how the increase of LOA from 4 to 7 in the Conflict Prediction & Resolution function changes the architecture.

20. Swihart et al. 1998.

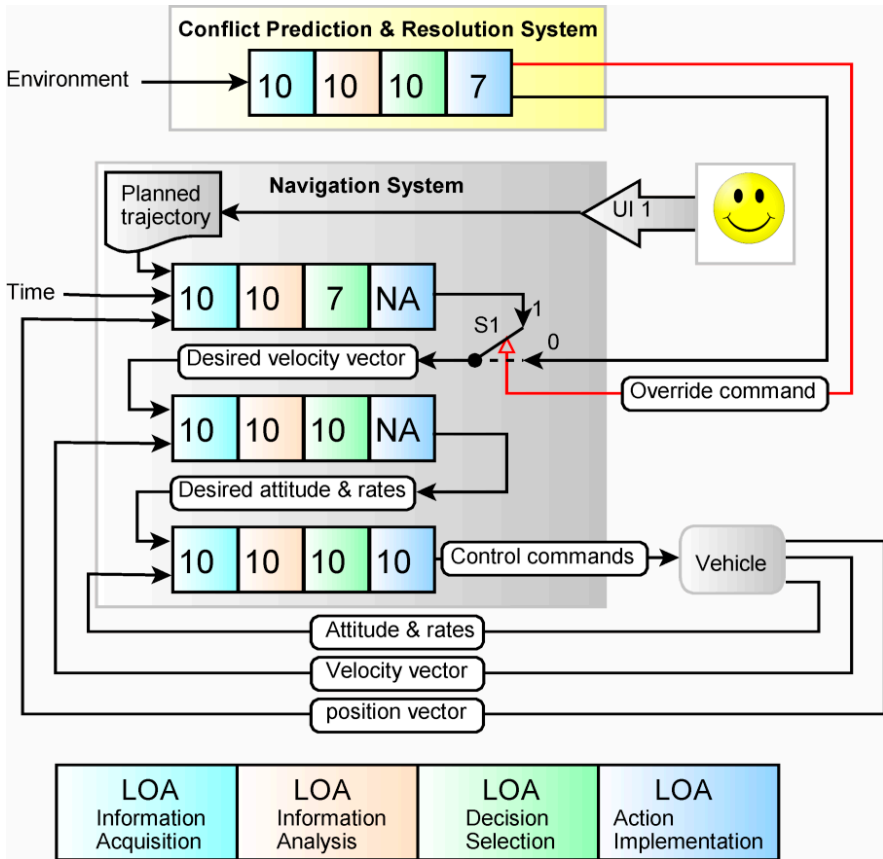


Figure 8.2. Increased Autonomy of the Navigation Function through Integration with a Conflict Prediction and Resolution Function

With the architecture illustrated in Figure 8.2 it is still possible to change the plan during a mission, but the system is now capable of executing this plan without requiring a human operator action to deal with unexpected traffic or ground collision hazards.

This example has illustrated how integrating existing automated functions eliminated the need for the human operator to close a particular control loop and made the system more autonomous. For a function in which the human operator is not an integrator but performs the four phases of information processing, this raises the question of what (performance) criteria a system

needs to meet in order to replace this function. From a design perspective there is a need for objective, quantitative performance criteria, but these do not always exist for functions performed by the human operator. For example, the requirements for self-separation (i.e., to remain well clear of other traffic) have not (yet)⁽²¹⁾ been quantified. The need for an analytical definition of ‘well clear’ is established on the prerogative that ‘if a technical system is to perform an equivalent function to pilots, it is necessary to have an unambiguous, implementable definition of the separation the system seeks to maintain with other aircraft.’⁽²²⁾

When trying to establish a basis for performance requirements for a function that is to be automated by analysing and quantifying how a human performs a certain task, there is a risk that alternative solutions are overlooked because the system design is too focused on replicating how humans perform the particular task. For safety-critical systems, performance-based approaches that include a Target Level of Safety (TLS) or Equivalent Level of Safety enable a more independent comparison.

Target Level of Safety

The required TLS can be achieved either through a combination of automated functions and a human operator, or by means of a fully automated function. In the application of the Required Navigation Performance concept to automatic landings,⁽²³⁾ the TLS is specified as 10-9 (the probability of a catastrophic accident per landing). In the so-called Category II configuration, the pilot has to decide (based on the visible location of the runway) whether to continue to land at or before a decision height of 100 ft. As such, the LOA can be classified as 6 (pilot can veto). In the risk analysis, the pilot is credited to realise a risk reduction of a factor of 10. As a result, the automated system is required to achieve a TLS of 10-8. This requirement is allocated to the continuity, accuracy, and integrity of the system. The configuration in which the LOA is increased to 7 can no longer take credit for the risk reduction provided by the pilot; thus

21. The SARP, a US government-funded organisation designed to bring together SAA community stakeholders to close known research gaps, suggested a quantitative definition for ‘well clear’ in August 2014. The FAA has suggested a modification in September 2014.

22. Weibel, Edwards, and Fernandes 2011.

23. Kelly and Davis 1994.

a TLS requirement of 10-9 applies to the automated system. In other words, to achieve the same TLS without taking credit for human operator involvement, in this example the system performance requirements in terms of continuity, accuracy, and integrity have increased by a factor of 10.

Alerting/Decision Thresholds

The basis for a conflict prediction system is the criterion that defines when a conflict exists. Based on the measured state, a prediction (e.g., through extrapolation) is performed to test whether the future (predicted) state meets this criterion. If so, a conflict is said to occur. Due to uncertainty (caused by sensor⁽²⁴⁾ and prediction inaccuracies and lack of intent information), there is a likelihood of missed detections and false alarms. For safety-critical systems, a primary requirement concerns the maximum allowed likelihood of missed detections. When taking uncertainty into account to limit the likelihood of missed detections, the likelihood of false alarms will increase. Because uncertainty increases with look-ahead time (the amount of time that is predicted ahead), the acceptable false alarm rate puts a limit on the look-ahead time for a conflict prediction function. The implication of a process in which the pilot is involved in decision making (e.g., remaining well clear) is that at a larger (temporal) distance from the predicted Closest Point of Approach (CPA) there is more uncertainty regarding the minimum distance that will be achieved at the CPA. Postponing a manoeuvre (in order to assess how the predicted minimum distance changes) will reduce the amount of unnecessary manoeuvres at the expense of more severe manoeuvring where a collision hazard exists.

Discussion

Many other examples of systems that exhibit autonomous behaviour can be found outside of the safety-critical applications. For example, there are technical ways (e.g., sensors, processing, and actuation) to increase the level of autonomy of the navigation. Software design processes that meet the high-reliability

24. An important factor contributing to uncertainty is the accuracy of the sensors used to measure the state that serves as the basis for the prediction. Hence, an increase in sensor accuracy typically allows the look-ahead time to be increased while continuing to meet the missed-detection and false-alarm requirements.

requirements⁽²⁵⁾ and fault-tolerant design and redundancy management techniques are also long-established engineering practice in this domain.⁽²⁶⁾

A major challenge concerns translating the performance that is (supposedly) achieved by humans into quantified system requirements. Overall performance is achieved as a result of skill, rule, and knowledge-based behaviour. For example, with safety-critical functions, an important reason for human operator involvement at the decision-making level concerns the merit of knowledge-based behaviour. Furthermore, humans' ability (skill) to detect patterns in noisy data can be exploited to reduce false alarm rates in ways that cannot (yet) be matched by fully autonomous (pattern recognition) systems. Attempts to replicate these specific human capabilities have yielded complex systems, and unanticipated failures are reported in interesting anecdotes. Yet approaches that enable a human operator to explore different (decision) options using the brute force capabilities of a computer to predict the outcome have been found to increase overall performance.

With a TLS-based approach, the contribution of a human operator to an increase in safety can be quantified as an incident-to-accident ratio for the associated function. To achieve the required TLS, an understanding of the potential failures is required. Wiener and Curry⁽²⁷⁾ distinguish between the following seven categories of automation failures:

1. failure of automatic equipment;
2. automation-induced error compounded by crew error;
3. crew error in equipment setup;
4. crew response to a false alarm;
5. failure to heed automatic alarm;
6. failure to monitor; and
7. loss of proficiency.

Categories 2 to 7 only occur for functions with a LOA of 7 or lower. It is emphasised that 'the assumption that automation can eliminate human error

25. RTCA 2011.

26. NATO 1980, 1990.

27. Wiener and Curry 1980.

must be questioned',⁽²⁸⁾ which introduces an important design decision. For functions with an LOA of 8 or higher, no credit for risk reduction can be taken for human operator intervention to achieve the TLS, thus increasing the performance requirements that the function has to meet. Also, this requires considering a potential (partial) transfer of liability from operator to manufacturer. Given that no human operator is involved, the required processes can be simulated faster than real time, providing the possibility to generate large datasets required to demonstrate compliance with a stringent TLS. For a LOA of 7 or lower, the designer will have to demonstrate how failures of categories 2 to 7 are addressed in such a way that the TLS is achieved. This will require human-in-the-loop evaluations that cannot be performed faster than real time.

Increasing Function Autonomy

There may be a multitude of reasons to require a certain level of autonomy, ranging from a reduction in operator workload to the need for (a more) independent operation of the system hosting the function. For unmanned systems, the available control link bandwidth, communication latency, lack of sufficient visual cues, and requirements concerning function availability in case of a control link failure will influence the minimum level of autonomy that can be selected.

With remotely operated systems, an important design decision concerns the path of the data flow that comprises the closure of the various loops involved in the navigation. Unmanned aircraft such as the Predator and Global Hawk have the capability to close the navigation, guidance, and control loop on board the vehicle. They can continue to execute the planned trajectory if the control link fails.⁽²⁹⁾ Olson and Wuennenberg describe how operator roles for unmanned aircraft change as vehicle autonomy increases: 'if the operator cannot intervene to affect a desired change in vehicle performance (e.g., if the operator is unable to change the rate of climb/descent to avoid a collision), then an intelligent automated detection/intervention system may be required'.⁽³⁰⁾

28. Ibid.

29. In case of a control link failure, the automated navigation system will select another trajectory, e.g., one that causes the aircraft to climb in an attempt to re-establish a connection that may be lost due to line-of-sight problems or the execution of a return to base trajectory.

30. Olson and Wuennenberg 2001.

However, ‘artificial or machine intelligence is quite different from human intelligence, and it serves no useful purpose to try to relate one to the other’.⁽³¹⁾ Rather than specifying ‘intelligence’ as part of the design requirement for a system that replaces one or more functions performed by a human operator, a specific TLS (or the requirement for an equivalent level of safety) should be the basis for the design, validation, and certification of functions that extend an automatic navigation system with the capability and authority to autonomously manoeuvre.

For example, current regulations to prevent collisions with other aircraft require the pilot to visually acquire potential intruders. Since unmanned aircraft do not currently have the means to (provably) meet this requirement, the workaround is to reduce the likelihood that an encounter with another aircraft occurs by temporarily prohibiting other aircraft from accessing the airspace in which the unmanned aircraft operates. This is comparable to how, with ILS, it is assured that no other aircraft are in the vicinity of the trajectory of the aircraft performing the automatic landing. However, this approach severely restricts the use of airspace.

Future regulations are likely to require a solution that provides an equivalent level of safety.⁽³²⁾ In this context, two sequential safety nets are envisioned. The final safety net is intended to be realised by means of an autonomous collision avoidance function. Autonomy is required to ensure function availability and continuity (i.e., to prevent a collision in a partly unpredictable situation) in case the link is lost. To stay below a given missed detection rate – while minimising unnecessary manoeuvres caused by false alarms – this system operates in a short time horizon (less than a minute). Given the requirement to achieve an equivalent level of safety as pilot visual identification of intruders to the airspace, the performance requirements for the autonomous collision avoidance function can be relaxed if the encounter probability is reduced. To achieve this, an additional threshold, the so-called well-clear boundary, is introduced. To help the pilot remain well clear, several concepts are being pursued for the conflict prediction and resolution function. These concepts differ in the design of the graphical user interface and in the LOA applied

31. Statler 1993.

32. FAA 2013.

for the Decision Selection block. Figure 8.3 illustrates the differences in LOA between the two safety nets, and the range of options for the Decision Selection block to provide information to remain well clear.

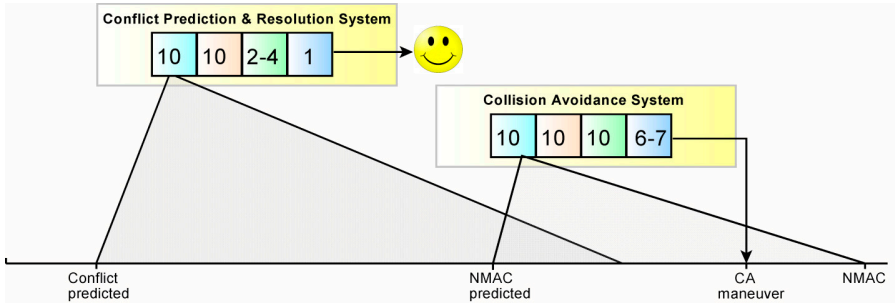


Figure 8.3. Two Sequential Safety Nets with Different LOA for Decision Selection and Action Implementation

The LOA of the Decision Selection block in the Conflict Prediction & Resolution system can be selected between 2 and 4. At 2, the computer offers a complete set of decision/action alternatives. For example, the conflict probe display⁽³³⁾ system computes *whether* (and, if so, *where*) a well-clear violation is predicted to occur for a specified range of manoeuvre options, and presents the resulting conflict space to the pilot. Every manoeuvre that is predicted to avoid the conflict space is a potential option to remain well clear. This is an example of the above-mentioned concept in which a computer is used in a brute-force approach to compute the outcome of a range of different inputs (a task well beyond the capabilities of a human); the pilot, taking into account all relevant information regarding the current situation, selects the most suitable one. At an LOA of 4, the system will only present a single manoeuvre option to the pilot.

The look-ahead time of the function that predicts whether a well-clear violation will occur is larger than that which predicts whether a *collision* will occur. The resulting increase in uncertainty will cause a higher number of false alarms generated by the Information Analysis block in the Conflict Prediction function. However, because the LOA of the Decision Selection block is limited

33. Theunissen, Suarez, and de Haag 2013.

to 4, a manoeuvre to remain well clear will only be performed if the human operator has decided that such a manoeuvre is warranted. Communication with air traffic control and/or the pilot of the intruding aircraft and information about intent are factors that will influence the pilot's decision. The fact that uncertainty reduces as the time to the predicted conflict decreases provides the pilot with the option to wait and obtain more information. An important requirement for the pilot-vehicle interface is that it provides the information needed to make the pilot aware of the uncertainty.

A well-designed pilot-vehicle interface should make missed detections (of a well-clear violation) extremely unlikely, while the information about context enables the pilot to ensure that the false alarms generated by the Information Analysis block do not lead to unnecessary manoeuvres. The result is that between the time that a conflict is first predicted to occur and the time that a collision avoidance system predicts a collision will occur, a significant amount of the encounters can be resolved by using the information presented to manoeuvre well clear. Thus, the encounter probability at the input of the Collision Avoidance function is reduced.

However, collision hazard geometries can develop within the look-ahead time used by the conflict prediction function. This may create situations in which the time to assess the situation is insufficient for adequate decision making, causing the collision avoidance function to autonomously execute a manoeuvre.

Discussion

How does the approach in the example differ from some of the other automation intended to support the human operator? Because they have a sequential approach, both systems have a high LOA for Information Acquisition and Information Analysis. For the conflict prediction and resolution function, this is needed to minimise the cognitive effort required by the human operator to understand the current situation and answer the 'what if' question regarding the future. For the collision avoidance function, it is needed because of the required high LOA for Action Implementation.

The conflict prediction and resolution function has a low LOA for the Decision Selection and Action Implementation block. Compared to often-mentioned automation issues (e.g., 'what is IT doing?', 'what is IT going to do?'),

and ‘Why?’) in which the concept relies on the depiction of integrated status information, the pilot is not required to understand *why* the automation has made a decision. He or she is only expected to make an informed decision based on the provided data. As with all automation interaction, this imposes requirements on the pilot vehicle interface that must be met through good design practices. However, it prevents the issue that fundamentally plagued the PA programme. The tendency of the pilots not to trust the decision made by automation in critical situations if it differed from what he or she expected. The decision threshold used in the collision avoidance function, which has the autonomy to command a manoeuvre without requiring pilot consent, must represent a situation in which it can be assumed with high confidence that immediate action is needed to prevent a catastrophic event, and the lack of action from the pilot indicates that either (1) a control-link failure has occurred or (2) that the pilot has either failed to detect the event or failed to decide in a timely fashion on a manoeuvre. Similar to the very final phase of an automated landing, once the probability of an accident due to unwarranted intervention in an autonomously initiated manoeuvre exceeds the probability that the continued execution will lead to an accident, one may wish to consider ‘locking out’ the pilot’s ability to override the automation.

A recent example of how the level of automation can play a key role in an emerging technology is related to detect and avoid (DAA), or the ability of remotely piloted aircraft systems to fulfil the ‘see and avoid’ requirements of today’s manned aviation. In the United States, the Radio Technical Commission for Aeronautics has established a Special Committee, SC-228, to develop minimum operational performance standards for a DAA system. The committee seeks to describe the levels of automation that the functions of a DAA system could obtain, and systematically categorise those levels into groups that form coherent possible systems. In working to develop a system that can (at least partially) replace the pilot’s ability to look out of the window and avoid other aircraft, a key point of contention is whether an algorithm is required to perform the task. For example, it may be obvious to designers that a radar should automatically detect a target, because there is no way for the human pilot to look at raw radar returns and distinguish positive from negative returns. On the other hand, it may be prohibitively difficult to develop an algorithm that can handle all of the possible encounter geometries and circumstances that a

pilot may routinely face while flying in complex airspace, but to the human pilot these conflicts are easy to resolve. By decomposing the functions and characterising the range of automation levels that would be appropriate for each, the group was able to constructively discuss a complex topic that had previously been a stumbling block. The result was a clear path that researchers had to take to answer more direct questions in order to develop an answer to the bigger problem. Work in this area is ongoing, but a committee white paper concluded that the human pilot does not need an algorithm to recommend a single manoeuvre to resolve a conflict (LOA 4); in the timescales of the self-separation function of DAA, the pilot could be presented with information (e.g., traffic symbology on a display) and allowed to make decisions that best suit the encounter and the mission.

Conclusion

Unlike autonomy designed for systems to play games, a failure of automated and autonomous navigation system functions can be severe or even catastrophic. Therefore, although an autonomous system may be technically feasible, the design process of each automated function must consider the implications of transferring authority from the human to the system. In this context, it is interesting to note that when the topic of autonomous airborne systems and the role of the human was being addressed in 1993, the conclusion was that ‘Human intelligence and the ability it confers to exercise judgment and, thus, to deal with unexpected situations will warrant the services of the human member in future systems.’⁽³⁴⁾ Although since that time developments in AI and available computer power have significantly increased, this conclusion is still as valid as it was over 20 years ago. Likewise, a 2003 US Air Force Scientific Advisory Board report stated: ‘[m]ission management, vehicle autonomy and human-computer interface have a long way to go’ and ‘[w]ith UAVs, like with manned platforms, the optimum solution is people and machines working together.’⁽³⁵⁾

Auto-GCAS and automatic ACAS represent systems that contribute to overall safety by having sufficient authority in the time-critical domain to

34. Statler 1993.

35. SAB 2003.

autonomously intervene in the navigation process. The rationale behind this automatic transfer of execution authority follows from the assumption that, due to the predicted high probability of an accident if no intervention is performed, the contribution of the human to the goal of safe navigation is less than that of the system.

However, outside the time-critical domain, in spite of continued advances in AI (and although technically, many envisioned types of navigation functions can be designed to operate autonomously), there are still many advantages to having a human operator involved.⁽³⁶⁾ Approaches that enable a human operator to explore different (decision) options using the brute force capabilities of a computer to predict the outcome seem a good option, which combines the advantages of the computer-based enablers of autonomy with the (still) unique human capabilities that justify the assignment of authority.

Clearly there are many design challenges concerning which human-machine interfaces best support the intended cooperation between the human and a system that comprises both automated and autonomous functions, and there is a vast body of research to aid in the design process. Furthermore, the increased use of automation to reduce dependency on the human operator for specific functions requires translation of the performance (supposedly) achieved by humans into quantified system requirements.

With the current trend of goal-based function integration and transfer of authority, the challenge is to apply the appropriate constraints in the transfer of authority from human to system outside the time-critical domain, and to identify and design the autonomy needed for those situations in the time-critical domain where an automatic transfer of authority to the system will save the day.

Implications beyond Autonomous Navigation Systems

This chapter started with the statement that, from an observer's perspective, it is difficult to distinguish between automated and autonomous systems. Nowadays, (mis)perception concerning autonomy in relation to unmanned systems that can be armed has resulted in (understandably) strong opinions about 'killer drones'. Knoop concludes that '[w]ithout revisiting the implications of robotic UAVs, with the capability of autonomously making seek-and-destroy

36. Statler 1993; SAB 2003.

decisions, as well as the legal and ethical challenges culminating from these technological advances, the future of UAVs seems prone to abuse'.⁽³⁷⁾

The existence of armed unmanned systems has also produced discussions about the question of meaningful human control. This chapter has illustrated that, from an observer perspective, the assumptions that form the basis of such a discussion can be fundamentally flawed. A misunderstanding of autonomy can yield a misperception of differences with today's rules of engagement. When addressing meaningful human control from a design perspective, the foremost question should be whether, *compared to the existing rules of engagement, there really is an intention to transfer authority over functions that have severe consequences* (e.g., an autonomous decision to use a weapon). If this is truly the case, in addition to addressing the resulting legislative questions, objective decision thresholds and performance requirements will need to be established, and the (transfer of) liability will have to be addressed. If this is not the case, a problem is being addressed that does not exist.

37. Knoop 2012. A detailed analysis of autonomy in relation to weapon systems is provided in Scharre and Horowitz 2015.

References

- Andes, R.C., and Rouse, W.B. *Specification of Adaptive Aiding Systems*. Report NAWCADWAR-92034-60. Warminster, PA: Naval Air Warfare Center, 1992.
- Berning, S.L., and Glasson, D.P. *Adaptive Tactical Navigation Program*. AGARD Conference Proceedings CP-499, 1990.
- Billings, C.E. 'Toward a Human-Centered Aircraft Automation Philosophy'. *The International Journal of Aviation Psychology* 1/4 (1991):261–70.
- Boyd, J. *Patterns of Conflict*. Briefing, 1986 (<http://www.ausairpower.net/JRB/poc.pdf>)
- Charnley, J. 'Navigation Aids to Aircraft All-weather Landing'. *The Journal of Navigation* 42/2 (1989):161–85.
- Clough, B.T. 'Metrics, Schmetrics! How do you Track a UAV's Autonomy'. Paper presented at the 1st Technical Conference and Workshop on Unmanned Aerospace Vehicles, Portsmouth, VA, 20—23 May 2002.
- Corrigan, J.D., and Keller, K.J. 'Pilot's Associate: An Inflight Mission Planning Application'. Paper presented at the AIAA Guidance, Navigation and Control Conference, 14–16 August 1989, Boston, MA.
- Emerson, T.J. and J.M. Reising. 'The Effect of Adaptive Function Allocation on the Cockpit Design Paradigm'. Paper presented at the 6th International Symposium on Aviation Psychology, 29 April – 2 May 1991.
- Federal Aviation Administration (FAA). *Sense and Avoid (SAA) for Unmanned Aircraft Systems (UAS) – Second Caucus Workshop Report*, 2013.
- Kelly, R.J., and Davis, J.M. 'Required Navigation Performance (RNP) for Precision Approach and Landing with GNSS Application'. *Journal of The Institute of Navigation* 41/1 (1994):1–30.
- Knoops, G.J. 'Legal, Political and Ethical Dimensions of UAV Warfare under International Law'. *Militair Rechtelijk Tijdschrift* 4 (2012):174–92.
- National Aeronautics and Space Administration (NASA). *Technical Note D-7843. Description and Flight Test Results of the NASA F-8 Digital Fly-By-Wire Control System*, 1974.

- North Atlantic Treaty Organization (NATO). *Fault Tolerance Design and Redundancy Management Techniques*. AGARD-LS-109. 1980.
- North Atlantic Treaty Organization (NATO). *Fault Tolerant Design Concepts for Highly Integrated Flight Critical Guidance and Control Systems*. AGARD-CP-456. 1990.
- North Atlantic Treaty Organization (NATO). *Knowledge Based System Applications for Guidance and Control*. AGARD-CP-474. 1991.
- Olson, W.A., and Wuennenberg, M.G. (2001). *Autonomy Based Human-Vehicle Interface Standards for Remotely Operated Aircraft*. Paper presented at the 20th Digital Avionics Systems Conference, 14-18 October 2001, Daytona Beach, FL.
- Palmer, M.T., Rogers, W.H., Press, H.N., Latorella, K.A., and Abbot, T. *A Crew-Centered Flight Deck Design Philosophy for High-Speed Civil Transport (HSCT) Aircraft*. NASA TM 109171. 1995.
- Parasuraman, R., Sheridan, T.B., and Wickens, C.D. 'A Model for Types and Levels of Human Interaction with Automation'. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 30/3 (2000):286–97.
- RTCA. DO-178C. *Software Considerations in Airborne Systems and Equipment Certification*. 2011.
- Scharre, P., and Horowitz, M.C. *An Introduction to Autonomy in Weapon Systems*. Washington, DC: Center for a New American Security, 2015.
- Schutte, P.C. et al. *The Naturalistic Flight Deck System: An Integrated System Concept for Improved Single-Pilot Operations*. NASA TM 2007-215090. 2007.
- Sheridan, T.B., and Verplank, W.L. *Human and Computer Control of Undersea Teleoperators*. Cambridge, MA: Massachusetts Institute of Technology, 1978.
- Statler, I.C. 'Technical Evaluation Report'. In *Combat Automation for Airborne Weapon Systems: Man/Machine Interface Trends and Technologies*. AGARD CP-520, pp. T1-T19. 1993.
- Swihart, D. et al. *Results of a Joint US/Swedish Auto Ground Collision Avoidance System Program*. Paper presented at the 21st ICAS Congress, 13–18 September 1998, Melbourne, Australia.

- Theunissen, E., Suarez, B., and Uijt de Haag, M. *From Spatial Conflict Probes to Spatial/Temporal Conflict Probes: Why and How*. Paper presented at the 32nd Digital Avionics Systems Conference, Syracuse, NY, 6-10 October 2013.
- US Air Force Scientific Advisory Board. *Unmanned Aerial Vehicles in Perspective: Effects, Capabilities, and Technologies*. SAB-TR-03-01. 2003.
- Weibel, R.E., Edwards, M.W.M., and Fernandes, C.S. *Establishing a Risk-Based Separation Standard for Unmanned Aircraft Self Separation*. Paper presented at the Ninth USA/Europe Air Traffic Management Research & Development Seminar, 14-17 June 2011, Berlin, Germany.
- Wiener, E.L., and Curry, R.E. *Flight-Deck Automation: Promises and Problems*. NASA TM 81026. 1980.

Auditable Policies for Autonomous Systems (Decisional Forensics)

Thomas Keeley

Abstract

This chapter will differentiate between the capabilities of human warfighters (today) and those of autonomous systems (tomorrow) and consider what needs to happen to enable the transition. Using judgment and reasoning, a human can evaluate situations and balance alternatives to make command and control decisions. We “trust” that the humans have had sufficient training to perform as expected. Humans are very reactive / adaptive systems. They can operate independently or as part of a team. They can take directions from above and can continue to operate when communications with command is lost. A “machine” deals with explicit numeric variables. For an autonomous machine to perform as expected, it will need to be programmed with judgment and reasoning skills that are equally as good—or better than—its human counterpart. The use of mass produced “machines” demands that the decisions and actions of these “machines” can be audited and “fixed” if unacceptable behaviour is detected. Policies controlling the behaviour of autonomous systems must be traceable to those creating the policies so they can be held accountable. This chapter will discuss some of the challenges that exist. It will also identify some specific tasks that need to be addressed and offer some options for consideration.

Introduction

Systems with autonomous capabilities are at the brink of mass production and mass deployment. This opens the door to large-scale problems if systems, or the policies they execute, make bad decisions or expose weaknesses.

The term ‘policy’ is used to define guidelines that are given to a machine to interpret information and control actions. This concept is suggested as a subtle distinction between programming rules to solve specific problems. It is similar

to policies and guidelines provided to humans through rules of engagement, laws and training. The terminology is expressed in more abstract terms that allow them to be applied to multiple situations. The assumption is that it will not be economically or technically feasible to write conventional IF THEN ELSE 'code' to address *all* of the problems an autonomous system will encounter.

Unlike systems controlled by individual humans who may make intermittent errors in judgement or reasoning, or just have a bad day, mass-produced machines will all operate the same way. So if there is a problem with the way the policy was encoded in the machine, that problem will be duplicated in every machine that is operating with the same policy in a similar situation. This is similar to a flaw that is exploited in Microsoft Windows: everyone who uses Windows is subject to that weakness.

People trust that humans have had sufficient training to perform as desired. Humans are very reactive/adaptive systems. They can operate independently or as part of a team, and they can take directions from above and continue to operate when communications with the executive command are lost. A machine, by contrast, deals with explicit numeric variables. For an autonomous machine to perform as expected, it will need to exercise judgement and reasoning skills as well as, or better than, its human counterpart. The use of mass-produced machines demands that the decisions and actions of these machines can be audited and 'rapidly fixed' if unacceptable behaviour is detected. Policies that control the behaviour of the autonomous systems must be traceable to those creating the policies so they can be held accountable.

This chapter discusses auditability and traceability, which are important for two primary reasons: (1) auditability, because you cannot fix what you cannot understand and (2) traceability, because humans are ultimately responsible for the ethical and legal behaviour of the systems. So if responsibility is not traced to the appropriate person, there is little likelihood that issues will be addressed. A third related concern is economic: auditability and traceability can only be effectively accomplished if it is economically feasible.

This chapter will differentiate between the capabilities of human warfighters (today) and those of autonomous systems (tomorrow) and consider what needs to happen to enable the transition. Using judgment and reasoning, a human can evaluate situations and balance alternatives to make command and control

decisions. We “trust” that the humans have had sufficient training to perform as expected. Humans are very reactive / adaptive systems. They can operate independently or as part of a team. They can take directions from above and can continue to operate when communications with command is lost. A “machine” deals with explicit numeric variables. For an autonomous machine to perform as expected, it will need to be programmed with judgment and reasoning skills that are equally as good—or better than—its human counterpart. The use of mass produced “machines” demands that the decisions and actions of these “machines” can be audited and “fixed” if unacceptable behaviour is detected. Policies controlling the behaviour of autonomous systems must be traceable to those creating the policies so they can be held accountable. This chapter will discuss some of the challenges that exist. It will also identify some specific tasks that need to be addressed and offer some options for consideration.

Significance of Auditability and Traceability to Autonomous Systems

Early machines were primarily human amplifiers. They became tools to facilitate profitability through mass production and automation.⁽¹⁾ In the military domain, machines were used to amplify the capabilities of an individual warfighter. In these early cases there was a large human component involved in the delivery of force. The humans either operated the systems – thereby controlling their behaviour – or programmed the absolute functionality into the systems. The primary responsibility for the machine’s actions could easily be traced to the operator or manufacturer, because the machines took all their directions directly from the humans who controlled their behaviour (i.e., the behaviour of the machines was traceable directly to the human). When things did go wrong, the legal system was there to debate and assign responsibility to the appropriate human(s). The migration toward autonomous systems will transfer more functions and decisions to the machines themselves. Many of the control decisions that were made by humans in the past will be made by the machines (autonomously) or in collaboration with some humans (semi-autonomously) in the future.

1. See Considine 1983, 1792, 281 for definitions of ‘machine’ and ‘automation’, respectively.

Key to understanding the behaviour of an autonomous system is the concept of a 'value system'. As machines take on functions that were previously the jurisdiction of humans, the policy makers will effectively be giving the machines a value system. An example of work in the area of establishing a value system in the military domain is the method of determining collateral damage estimation (CDE) values for weapon selection in engagement decisions.⁽²⁾

To understand the problem, one only has to consider the types of decisions with which the autonomous system will be challenged. In the past, machines were built to handle specific tasks (if this, then do that). There was no concept that the machine had to know why or had to justify its actions; that responsibility was left to the operator. The operator's decision about when to press the start button was influenced by the need to balance the directive given by his or her manager with safety concerns, the need for rest or refreshment, and pressures to clear the workspace of other workers who might be harmed. The operator's value system weighed the pressures of pushing the start button vs. waiting to resolve other issues. Autonomous systems will have multiple ways in which to address multiple problems simultaneously. They will balance alternatives to determine the 'best' way to pursue their goals (i.e., tactical decision making). They will also address multiple problems that will require them to distribute resources (e.g., allocate force) and perform duties that were done by humans in the past. These systems will simultaneously be considering short-term goals and long-term objectives. Such goals and objectives include engaging a target now versus surviving to fight again another day, balancing risk and reward (when and how), and considering the pros and cons of every action (should I, or shouldn't I?). The systems will need to decide between competing goals (tactical objectives, support friendly forces). They will plan, react to change, and revise their plans. They will adapt and self-organise. They will collaborate and organise to address problems they have never encountered before – which is a significant differentiator from past machines that were programmed to address a specific task. Thus autonomous machines will likely not have been tested against the exact characteristics of every task.

Change is another characteristic of the problem space that autonomous systems will be challenged with, especially in the military domain. Intelligent

2. CJCS 2012.

adversaries will change their tactics whenever they find that their current tactics are no longer valid. At the same time, new sensors and actuators will enter the marketplace to enhance and extend the capabilities of the autonomous systems. These changes will necessitate continued modifications to the autonomous system's cognitive models. Any changes will have a ripple effect through the development and audit systems, which will require that the cost and schedule impacts of development *and* audit are considered throughout the life cycle of the autonomous system.

To clarify, autonomous systems discussed in this chapter are considered to be bounded, adaptive machines that: (1) can be mass-produced with a static set of inputs and outputs, (2) can be assigned new goals, and (3) can adapt and react to those new goals with an understanding of their capabilities. These are *not* 'evolutionary' machines that can extend their input and output capabilities on their own, nor can they reproduce themselves automatically.

Human-controlled Machines versus Autonomous Machines

Human-controlled machines come with operating instructions: how to turn them on/off, how to adjust or tune them, and how to execute certain functions. The human operator is given guidelines (policies) for how to operate them, for example safety guidelines, operational procedures, rules of engagement, rules of law, preventative maintenance guidelines, etc. The human operators are trained in the operation of the machines (i.e., how to apply those safety guidelines, rules of engagement, and rules of law). The humans integrate this training into their own understanding of the language used in the training, which includes past experiences and prior training on morality and ethics. The machine builder ties the machine inputs (those used by the human operators and the sensors that may sense real-world states) to mechanical functions that have been programmed to perform the functions that the machines can execute. The human provides the judgement and reasoning functionality, and the machine builder provides the repetitive rules or programs in the human-machine partnership to pursue goals.

Similar to human-controlled machines, autonomous systems will be given policies that tell them 'how to think' when pursuing goals. These policies will tell the machines how to work with other machines and with humans on/in

the loop. While it may be conceptually possible to extend the current programming paradigm to transfer human judgement and reasoning capability into ‘programmes’ so that the machines can do everything humans do, this is not likely to be successful for economic and scheduling reasons.⁽³⁾

As machines process numbers, policies for autonomous machines need to be explicit. They cannot be subjectively interpreted, like policies for humans that are documented in the English language. Policies for humans depend on a consistent interpretation, yet every word is subject to a slightly different interpretation by every individual human. The policies for machines will define a ‘value system’ that will allow them to pursue goals on their own. The value system will be created by subject matter experts, just like the policy makers for humans, but they must be created in a manner that can be reduced to numbers and functional relationships. Because the policy, expressed in a mathematical form, will be executed in the same way by every identical machine, the machines can be mass-produced. And because the behaviour of any machine can be traced to a value system, this behaviour can be audited and traced back to the fundamental policies and information sources.

Auditing the Behaviour of a System

The term ‘audit’ comes from the Latin term *auditus*, the sense of hearing. Over time it has evolved to define an official examination and verification of accounts and records, especially financial accounts. More recently it has been used to refer to the inspection or examination of an entity in order to evaluate or improve its safety or efficiency, for example. The purpose of this chapter, however, is to focus on auditing the behaviour of autonomous systems (or machines) and to trace responsibility to the humans that defined that behaviour.

In human societies, human behaviour is taught through parental guidance, religious guidance, formal schooling, and peer interaction. Social feedback mechanisms attempt to correct or tune that behaviour model. Every human is unique, and each is exposed to different influences throughout his or her life. A large segment of the human population is involved in the feedback loop

3. The F-35 program has 8 million lines of code, yet it still requires a human pilot to make the judgement and reasoning decisions. See Lockheed Martin ND.

that attempts to influence the behaviour of other humans. The audience for auditing autonomous systems can be analysed from four different perspectives:

1. *Administration*. At the top level there are autonomous systems, with physical capabilities/tools/weapons. These are the machines and the machine resources that commanders use to pursue their goals. The responsible humans at this level choose/define the necessary functionality to accomplish a set of objectives, and decide which machines should be applied to which of these objectives. The humans also participate in the acquisition and deployment strategy, and are interested in ensuring that the autonomous systems at their disposal are doing the jobs they are expected to do, at the expected level of efficiency and cost. When things go wrong (or are perceived to go wrong) in the field, a legal audit may be required.
2. *Tactical decision making*. Each machine has policies that govern behaviour (guidelines that describe how to integrate valued information and distribute resources). The responsible humans at this level are the controllers or subject matter experts (SMEs) who tell the autonomous systems how to integrate information (how to 'think'). They provide the general guidance about how the autonomous systems are to address problems they encounter (embedded policy development and review). They also define how the autonomous systems will apply their capabilities toward objectives, according to the rules of engagement and international law, by creating policies that the machine can understand. These individuals specify the 'value system' of the autonomous system. They are also responsible for tactics and strategies. In an adversarial situation, the opposition will be adapting to changing tactics. So when tactics fail and goals are not achieved, it will be necessary to understand why, and adjust the tactics where appropriate (or analyse why the autonomous system was not able to adapt on its own).
3. *System architecture*. Each machine has real-time control capabilities (general operating system with rules that transform real-world data into values and provide specific methods of performing low-level activities). These capabilities might be termed the system infrastructure. The responsible humans at this level are the system architects who define how the system works. They are responsible for code reviews and delivering certifiable

code. These are also the engineers who know how to write the ‘control code’ that makes things happen.

4. *Sources of information.* There are information sources, sensors, and actuators (real-world inputs and outputs). At this level, raw data is transformed into information (mechanical review). The responsible humans are those who provide actionable data with a satisfactory confidence level. Auditors at this level are interested in ensuring that the humans and sensors provide the right information to the machines.

Designing for Auditability and Traceability

Before discussing the design and auditing process, it may be appropriate to consider the complexity and characteristics of the primary tasks and derived tasks that the autonomous system will be challenged to perform. For humans, some of these sub-tasks are never consciously thought about. But for an autonomous system, a human will have to provide guidance about how to address these tasks, as well as all other sub-tasks. In some cases these decisions and actions can be handled with conventional code. For example, if the autonomous system has an OFF button, the code to handle that could be as simple as IF OFFBUTTON THEN STOP.

But many of the autonomous system problems will be more complex. For example, a move is likely to be a sub-task. It might be assumed that an autonomous system can be given GPS coordinates and will just ‘go there’. Yet it will be in some state that might be qualified by its location, heading, fuel supply, weapons/tools, speed, etc. The autonomous system may also have to deal with other entities – either friendly or adversarial – each of which has its own characteristics. If the move is part of a larger task to move to a certain position and combine with others for a collaborative operation, the failure of any component will jeopardise the mission. While some of the startup and shutdown processes could be simple enough to handle with hard-coded solutions, others will have to adapt to the situation. Throughout the simple ‘move’, the autonomous system will have to recognise and adapt to change (i.e., to its health/fuel supply, changes in its environment or path). External intelligence may provide advisory information about risks or other opportunities that could cause it to re-evaluate its plans.

An example of a move command is moving from one side of a room to another, moving across the streets in the Middle East, moving through an environmental disaster area. The autonomous system would then have to do something, and to consider how to perform that task (which will likely have consequences). Using autonomous systems involves asking machines to perform tasks that were previously assigned to humans, so judgement and reasoning capabilities need to be transferred to these machines. A related issue is ensuring that the information is interpreted sufficiently to make a decision or take an action. Temporal data are part of decision making, as they help determine the importance of the information. The age of information also needs to be considered, since old information about one thing may be considered different than old information about another (e.g., the location of a building versus the location of a human target/friendly forces). Thus auditing the policies that control the behaviour of autonomous systems will involve some parties that are interested in the policies that define how the autonomous system moves from one point to another. It may be a completely different group that wants to know why an autonomous system engaged with a particular target, and a different group still that is interested in why the autonomous system decided to self-destruct (or partner with another group, or disengage from a target, or reject a command, etc.). All of these decisions will be based on the 'value system' imparted to the autonomous system, as discussed earlier.

Due to the complexity of the new policy development and processing tasks that will be challenging the autonomous systems, it is necessary to review of what one might want to ask of the autonomous system. This might be a design review, i.e., a preventative maintenance task, such as validating that a sensor provided the appropriate value to the cognitive process. It could also be a normal 'after mission review' or an event-specific review (because of the nature of the event). Essentially, an auditor wants to know if something is broken or incorrect in the autonomous system, or whether faulty information was provided to the autonomous system from a human or external information source.

During development, a significant effort will be made to ensure that the autonomous system will perform as desired. The first level of testing will likely be similar to testing and certifying conventional software systems, checking for boundary conditions, loops, system performance, etc. to eliminate crash situations. Testing how the system will solve problems will be a new challenge

(or at least expose many more conditions to test against). Experience suggests that this will be an iterative process of specification, testing, refining, and testing again.

Eventually the system will move to an emulation and simulation phase, in which one is evaluating events, and something happens or a decision is made and someone wants to know why. There may still be some complete policy reviews from the top down, but most will analyse events, and how they were addressed.

Because autonomous systems add complexity to the overall system, a critical aspect is cost. If auditing the system is more costly, it will be performed less often. This would introduce more problems into the field and/or delay the introduction of new tactics, sensors, or tools that could solve problems more effectively. Black Swan events⁽⁴⁾ have the potential to be very costly. This event review can be broken down into two levels:

1. What did the autonomous system consider when it did what it did?

Those interested in this question want to understand the basic inputs and outputs of the system. An external reviewer may be thinking ‘Why didn’t the autonomous system see those children playing next to that building before destroying it?’ The interested parties will know (by a process of elimination) what was not considered in the decision-making process. The auditor will be provided with the ‘what’ term in the type of event that occurred. This can be termed a ‘*high-level audit*’, in which the auditor/reviewer is looking primarily at which inputs were used to influence a decision or action. A common method of displaying a high-level view of relationships is the ‘influence diagram’. When the influence diagram (as shown in Figure 9.1) is automatically generated from the operational policy, the reviewers can be assured that they are getting the correct picture of the process, and the cost of creating the information model will be negligible. An interactive influence diagram will allow the reviewer to focus on specific parts of the overall policy. Figure 9.1 shows an example of interconnected bubbles, with arrows showing the direction

4. Taleb 2007.

of influence. Viewing interactive influence diagrams on a computer screen would expose the titles of the terms.

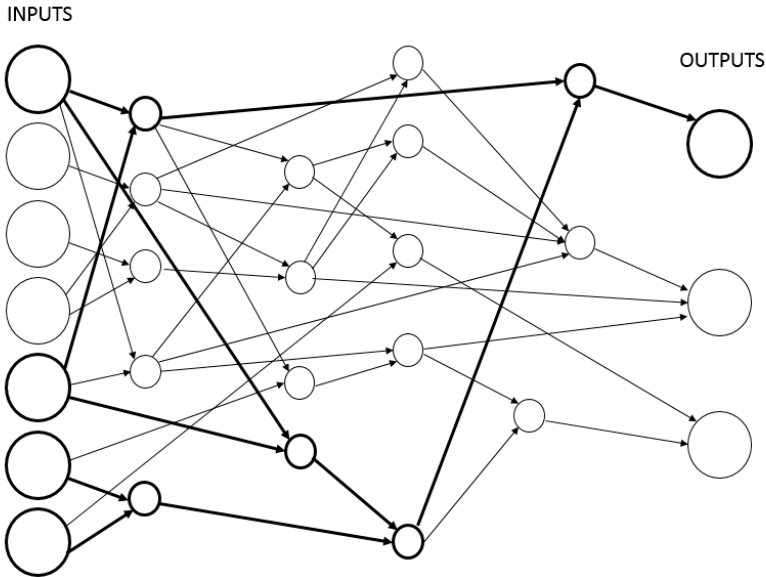


Figure 9.1. Sample Influence Diagram (Automatically Generated from Policy Source to Ensure Accuracy)

2. How were the inputs integrated to arrive at a decision or control action? Those interested in this question want to understand the details of how the inputs were integrated to influence the behaviour of the system when it made a decision or executed an action. This can be considered a *'detailed audit'*. Beyond 'what', the auditor will be interested in 'why/why not', 'when' (a specific decision or action was applied), 'how', 'how much', and potentially 'where' (or how the location for a decision or action was determined). The auditor will be interested in the age and the trustworthiness of information, and how each was determined. In some cases, a series of events (intermediate decisions and actions) may need to be reviewed to see what prior events led to subsequent decisions and actions. For example, consider a simple fuel gauge. The availability of

fuel will play a part in many decisions, because the fuel level will impact the ability of the autonomous system to accomplish its tasks. A human may periodically look at the fuel gauge in his/her car. As it gets to a half tank, the fuel gauge may be monitored more frequently, along with the perceived distance to a refuelling station. The decision to proceed without refuelling will be balanced against the time one needs to be at a certain point at a certain time, along with other factors. A curve defining the importance of fuel in the decision-making process is possibly exponential in shape. But the curve is not a constant because it could be influenced by damage to the drive train, new information about the availability of fuel, or the need to detour from the planned route. So how the fuel gauge is interpreted will be based on numerous influencing factors, each of which has more or less impact on how the fuel situation affects the overall situation. In this case one would be identifying the 'value' of fuel in the decision-making process. The auditor will be validating the value system of the entire autonomous system. Decisions and actions will be continuously updated in the autonomous system, because the importance of driving and blocking influences will change how the autonomous system interprets information. This means one is dealing with multiple inputs and multiple inter-related outputs that may be difficult to articulate. Some terms (distance to x, for example) may have very different impacts on different parts of the problem space. As one gets closer to x, one may get a better shot at a target. At the same time, one may be moving into a higher-risk area, and be using up more fuel and thus reducing the opportunity to get home. Figure 9.2 shows the use of graphical bars to indicate value. By displaying values graphically, it is easy to visualise the value system and see how valued information is accumulated. If one can 'see' how information is integrated, and change how it is integrated without translating concepts to formulas and onto code before testing the integration concepts, this would accelerate the development process significantly.

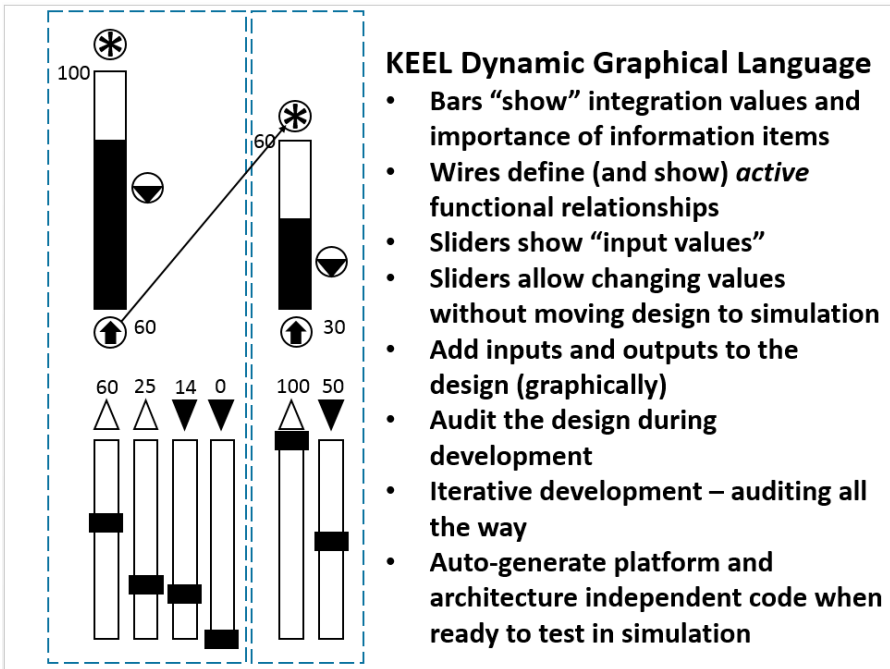


Figure 9.2. Dynamic Graphical Language Exposes the ‘Value System’

It will be important to consider the audience for the auditing process as well as economic considerations. Those auditing the policy-related behaviour are (in most cases) not the programmers. They are the individuals responsible for the behaviour of the system as it appears to the outside world: politicians, lawyers, acquisition specialists, sociologists, field commanders, policy makers, ethicists, and the general public. The programmers and engineers will want to audit the systems to confirm that if the policy says to do something, the command is executed correctly. They will want to make sure that a sensor that measures a physical parameter correctly transforms the real-world signal into the appropriate piece of information for further processing. The challenge is to create a cost-effective way to provide information in a form that can be understood by the policy auditors.

Defining and Executing Policies

Two approaches may be considered for defining and executing policies. A conventional iterative development process for complex systems might look something like this:

Approach A: Conventional Approach (treat the code for policies just like any other code)

1. The domain expert (policy maker) describes the policy objective in a text document and passes it on to the mathematician. Potentially execute a legal and safety review of the policy.
2. A mathematician translates the concept described in the text document into formulas.
3. A software engineer translates the formulas into modularised computer code.
4. The software engineer debugs the code and iterates until the code compiles successfully.
5. The software engineer writes a wrapper around the algorithm so it can be tested by the mathematician.
6. The mathematician refines the formula. Go back to step 3.
7. The software engineer documents the design (maybe).
8. The algorithm is integrated into a simulator for further testing by the domain expert. Potentially execute a legal and safety review.
9. The domain expert revises the description of the solution to more appropriately define the desired behaviour – go back to step 2. Potentially execute a legal and safety review before returning to step 2.
10. The problem changes – go back to step 1.

As one would expect, when problems become more complex, more commented lines of source code take longer to generate, test, debug, and evaluate. More complex policies will also likely take more passes through the system in order to get the policy to work as desired.⁽⁵⁾ Evaluation at this level involves checking that the code is working correctly, independently of the objective of the policy.

5. DOD 1985.

When ‘policies’ are developed, they may also require legal and safety reviews. These types of reviews focus on the objectives of the policy, and they assume that the code is correct. This results in the need for humans with other skills to participate in the evaluation and review process. These individuals may not be fluent in software. Using Approach A will require additional services to be integrated into the design to help this group of individuals understand what is going on, which adds more complexity to the overall solution. One common method is to integrate human-machine interface components into the system design that can expose some of the variables in a more human-friendly manner. While this may be helpful in the simpler cases, it may not show all the interactions between all the variables. It may not expose the ‘value system’ and ‘information fusion process’ to the auditor. If a detailed review is needed, it will be necessary to see all of the input variables *and* how they are integrated in order to check the validity of the policy.

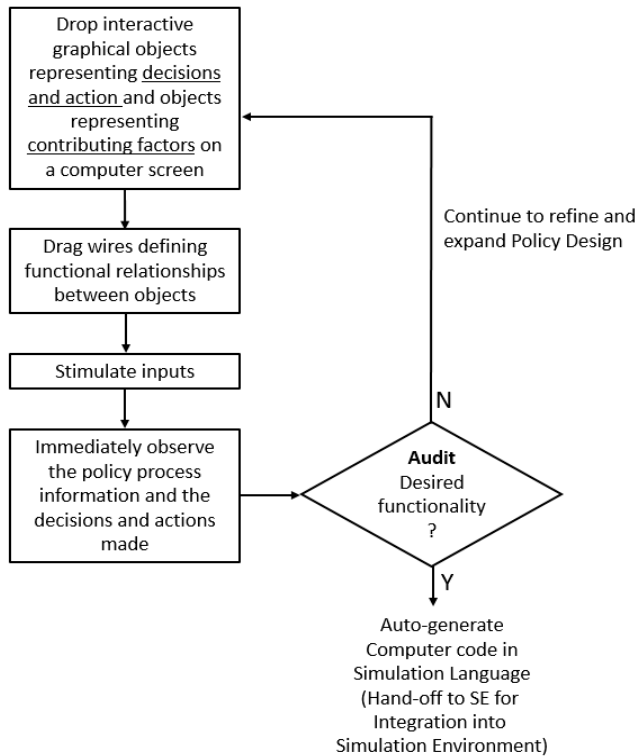
Approach B: Optimised Approach (Policies Define How To Think rather than How To Solve Specific Problems)

Using an optimised approach to define and execute policies will streamline the process. Using an interactive graphical policy definition language,⁽⁶⁾ the policy maker can develop and test the policy before it is ever translated into ‘code.’ Legal and safety reviews could also be accomplished at this level as long as the language was understandable with minimal training. A development environment that supports automatic code generation in multiple languages would allow working code to be tested in simulation with little extra effort.

1. The domain expert (policy maker) describes the policy objective using an interactive graphical policy definition language. The policy maker tests and revises the policy until it seems to perform as desired. Safety and legal reviews are performed using the interactive language to assist in testing. Once that level of testing is complete, code is auto-generated and handed to the software engineer. Potentially execute a legal and safety review. This step equates to steps 1 through 7 of the conventional approach (Approach A above).

6. Compsim, ‘About KEEL Technology’.

Dynamic Interactive Language for Policy Development



Operations performed by domain expert, without requiring support from mathematician or software engineer

Figure 9.3. Audits during the Policy Development Cycle

2. Working auto-generated cognitive/decision-making code is integrated into a simulator with the control software for further testing by the domain expert and tacticians. Behavioural auditing takes place where the system is tested in numerous scenarios. Decisions, tactics, and resource allocations are audited in realistic (dynamic) scenarios. Potentially another legal and safety review is performed.

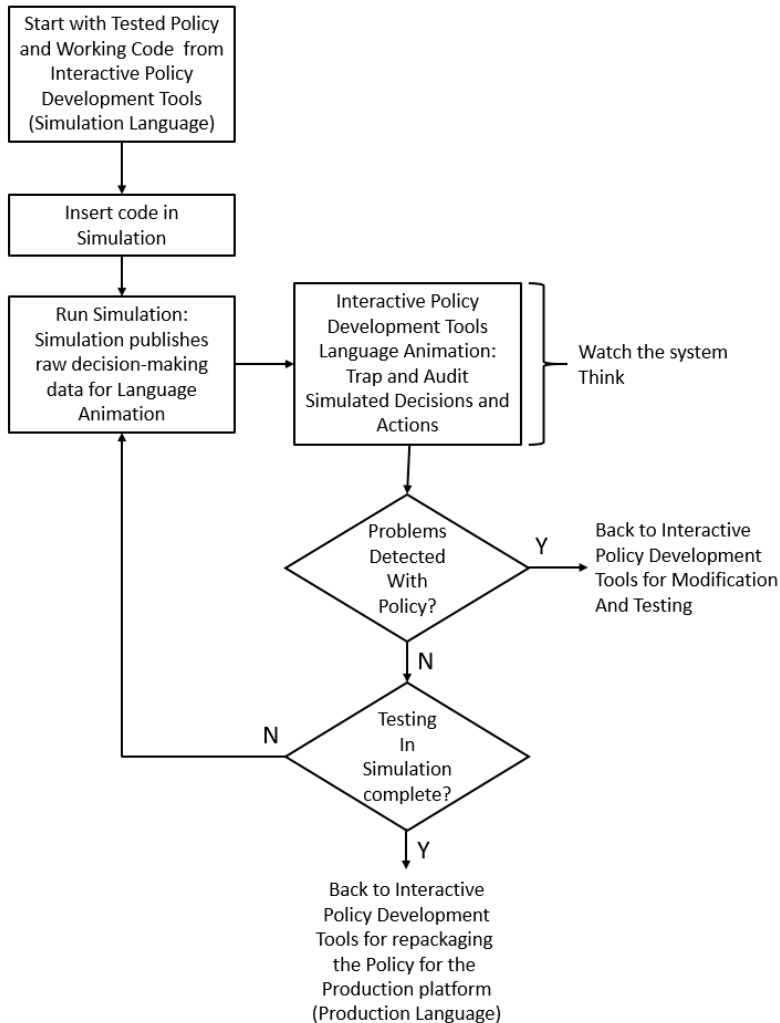


Figure 9.4. Audits and Policy Refinement During Simulation Testing

3. The domain expert revises the policy definition using the interactive graphical policy definition language to refine the desired behaviour. Go back to step 1. If satisfactory behaviour is exhibited in simulation, turn code to production (potentially in another programming language).
4. The problem changes – go back to step 1.

Allowing the policy maker to interact with his/her policy during the policy development process without having to translate from concept to formula, and again from formula to code, streamlines the policy development and review process. Using an interactive graphical language allows the policy maker to observe how the policy will operate in a dynamic environment and to refine the value system encapsulated in the policy in seconds. One can expect that a significant amount of testing of autonomous systems will take place in simulated environments before they are deployed in the field. It will be important to be able to observe how the autonomous system is integrating dynamically changing information in real time. A methodology that allows the tester and the policy maker to observe the information fusion process will be mandatory.

Auditing Policies versus Auditing Code

Auditing code in a high-level language can be compared to auditing machine language (1's and 0's). There is a general 'trust' that a high-level language (e.g., C++ or Java) can be compiled accurately into machine language. Almost no one today debugs system code at the machine-language level (except compiler manufacturers or microprocessor manufacturers). Instead, users 'trust' that the compiler accurately translates concepts encoded in IF|THEN|ELSE statements into the appropriate machine code.

To audit a program developed in a high-level programming language (Approach A above), the developers do code walkthroughs to review the logic. The developers put breakpoints in the software and step through critical routines. Then they instrument the code with other code to log interactive processes. Test code is then created to test boundary conditions. Other tools are utilised to test coverage to ensure that all branches within the code are tested. In a large system, this is a very costly exercise.⁽⁷⁾

Depending on the development process used to capture the policies of the autonomous system, the cost of auditing the behaviour of autonomous systems could increase exponentially due to the complexity of the tasks that the autonomous system will be performing. Yet if the policy development process uses a methodology that does not require translation to formulas and again to code before the policy can be audited (Approach B above), this will save considerable

7. Open Web Application Security Project ND.

time and resources because the people auditing the behaviour of the system can ‘see’ how the machine is interpreting the information without translation or subjective interpretation of values. The policy maker *and* the auditors can ‘see the value system in operation’ and understand the same information and value system that the autonomous system sees, and recognises.

The Need for Speed

Once autonomous systems are fielded for production, one hopes that the code has gone through an acceptable safety review.⁽⁸⁾ One also assumes that the policy has gone through both a conceptual and legal review, and a performance review using simulation. Similar to the policy review in simulations, there will be formal after action reviews.⁽⁹⁾ Since the behaviour of autonomous systems (machines) will be dictated by ‘valued information’, and by how that information is integrated to make decisions and control actions, it should be possible to determine the root cause of any decision or action (from the autonomous system standpoint). In other words, all information can be traced to its source.

Just because it will be ‘possible’, however, does not necessarily mean that it will be easy. If Approach A is used, then an after-mission review will revert to a code review or a review of recorded information presented on a display that is removed from the actual policy. If Approach A is used to develop the executable policy, it will be much more difficult to understand why the autonomous system did what it did (or did not do what it was expected to do). Yet if Approach B is used, it will be possible to examine in depth what happened in a very short time without undue effort. An after action review of critical decisions made by an autonomous system should be possible in less than 30 minutes.⁽¹⁰⁾

The Audience for Auditability and Traceability

Because machines cannot be held responsible for their decisions and actions, a number of groups will be interested in the behaviour of the machines. Each group will have its own focus of interest, and will want to audit the behaviour of the machines so they can understand and predict why they do what they do

8. NAVSEA ND.

9. Susanne Salem-Schatz, Ordin, and Mittman 2010.

10. Compsim, ‘Auditing a KEEL-based Policy’.

(proactive audit). After the fact, they will be interested in reviewing why the machines did what they did (reactive audit). Other groups will be interested in assigning/recognising responsibility for the behaviour of the machines. Some of these machines, even if they have autonomous characteristics, will collaborate with humans and other systems. The humans may provide information that is fused with other sensed information within the machine and thus influence its behaviour. In this case, being able to trace the information to its source will be of specific interest to groups that want to assign responsibility.

Table 9.1. Parties Interested in Auditability and Traceability of Autonomous Systems

Interested Party	High-level Audit	Detailed Audit	Traceability
Government / Regulatory Organisations	●		●
Policy ‘Maker’ (SME)	●	●	●
Safety Organisations	●	● ¹	●
International Humanitarian Organisations	● ¹	●	● ¹
Lawyers	●	●	●
Operations/Acquisitions/ Product Owners	●	● ¹	●
System Engineers/Architects (Product Developers)	●	● ²	
Software Engineers (Developers)	●	● ²	
Field Commanders	●	● ¹	●
Civilian Population in Countries / Organisations Using AxS	● ¹	● ¹	● ¹
Civilians/Entities on Receiving end of AxS Actions	●	● ¹	●

1 Indicates interest sometimes, but probably not all the time

2 Primarily during development and debugGovernment/Regulatory Organisations

Government and regulatory organisations are interested in governing how their positions are translated into actionable policies. When things go wrong, they will be interested in tracing the behaviour to its source, especially when they are pressured by the international community or the general public. Initial audits by government and regularity organisations will likely focus on high-level reviews of the policy. Subsequent reviews will likely focus on traceability in order to assign responsibility. Governments will also be concerned with

international relations, so the political ramifications of the use of autonomous systems will be of significant interest. Governments will want to highlight how the public can be assured that every effort is being made to ensure that the autonomous systems are safe and effective tools of the modern era. The international community will also debate the fact that some autonomous systems will be fielded that are not auditable (and the decisions and actions of which are not traceable). Compliance with international law might be questioned if systems were released that could not be audited.

Policy ‘Maker’

The policy maker is the SME who creates the policy. In many cases these individuals have not been challenged with creating ‘mathematically explicit’ policies before. So it is likely that the policy development phase will be an iterative process that results in a ‘need for speed’. The policy maker will also need internal access to the details of the policy so they can see and interact with the ‘value system’. For them, the high-level audit will primarily be used as a tool to explain the policy to others. If/when something goes wrong, they will need to be able to trace any issues to the root cause so they can be rapidly fixed. These individuals will also be responsible for mapping new data sources into the policy. This process is likely to repeat itself throughout the life of the autonomous system. If/when things go wrong in the field, the policy maker will want to trace the issue back to the policy, to driving information sources, and to any humans that might have had something to do with the information used by the autonomous system. Since the policy maker will ultimately be responsible for the functionality of the policy, this individual can be expected to repeatedly review, refine, and extend it. The level of effort that the policy maker requires has a major life-cycle cost impact on the overall success of the autonomous system. The policy maker will definitely be interested in traceability, because he/she will be the target of an inquiry.

Safety Organisations

Safety organisations will be interested in the quality of both the policy and of the code implementing the policy. Safety organisations are always under cost pressures as they ensure that systems are safe. So any mechanisms that simplify their work have an economic impact on the life-cycle cost of the

autonomous system. Also, because it is likely that the embedded policies of the autonomous systems will continue to evolve, their audits will be recurring. The safety organisations will be primarily interested in performing detailed reviews, both before and after deployment.

International Humanitarian Organisations

International humanitarian organisations will primarily be interested in the behaviour of autonomous systems after deployment, especially when policies impact target populations. They are likely to be interested in the embedded ‘value system’ that defines how the autonomous system values human life and the goals assigned to the system. It is likely that these organisations will not be given access to a detailed audit.

Lawyers

Lawyers will likely be involved before an autonomous system is released to production, and after it is released if/when something goes wrong. Before release, they will be interested in validating that the policy is consistent with international law and the standing rules of engagement. If they are able to review the policies without doing IF | THEN | ELSE code review, their review process will be much easier. Lawyers will be interested in how the policy maker captured the value system for the autonomous system, and how goals assigned to the system capture terms like ‘anticipated military advantage’ in a mathematically explicit manner. If/when things go wrong after deployment, the lawyers will be interested in reviewing the policy and tracing any problems to the root cause (i.e., to the policy itself, or to information sources including faulty hardware or contributions from humans in the loop). If an audit can be accomplished cost effectively and in a timely manner, this can reduce the life-cycle cost of the system.

Operations/Acquisitions/Product Owners

This group is interested in auditability and traceability, and they are the primary decision makers in selecting and using autonomous systems. They want to benefit from all of the capabilities of these systems so that more can be done with less effort. In most cases a high-level audit will be sufficient for this group, which will be interested in traceability issues if/when things go wrong so that they can modify their acquisition strategy.

System Engineers/Architects (Product Developers)

The engineers and architects are interested in the overall performance of the systems they are designing. They are primarily interested in reviewing the policies before the systems are placed into production, and in architectural considerations that determine how the policies are integrated into the main control loop. This group is interested in connecting the information sources to the policy and controlling the behaviour of the autonomous system based on the embedded policy. A high-level audit maps the physical world to the policy.

Software Engineers (Developers)

The software engineers are responsible for integrating the policy into the system architecture. This assumes that there is a separation between the policy development, the packaging (created by the policy maker), and the actual coding of the policy (Approach B above). In this case, the software engineers will primarily be interested in a high-level audit to ensure they have correctly integrated the information sources into the policy. However, if the software engineers are involved with manually translating the concepts of the policy maker into code, then they will be heavily involved in any kind of audit and will be involved in root cause traceability if/when things go wrong.

Field Commanders

In the field, these are the individuals responsible for deploying the autonomous systems on missions. They will assign the goals and provide mission-specific information that will guide the autonomous systems in the pursuit of those goals. They will be interested in after action reviews to validate the entire system's performance. They will be interested in traceability, especially if the performance might be traced back to pieces of information that they provided

for specific missions (e.g., anticipated military advantage, target value). One can expect that adversaries will attempt to use cyber warfare tactics to trick the autonomous systems. The field commanders may be the first to detect tricked behaviour in after action reviews, and they will want to advise others as soon as possible so problems can be fixed. Figure 9.5 shows the detection of a problem and going back to the drawing board to fix the policy.

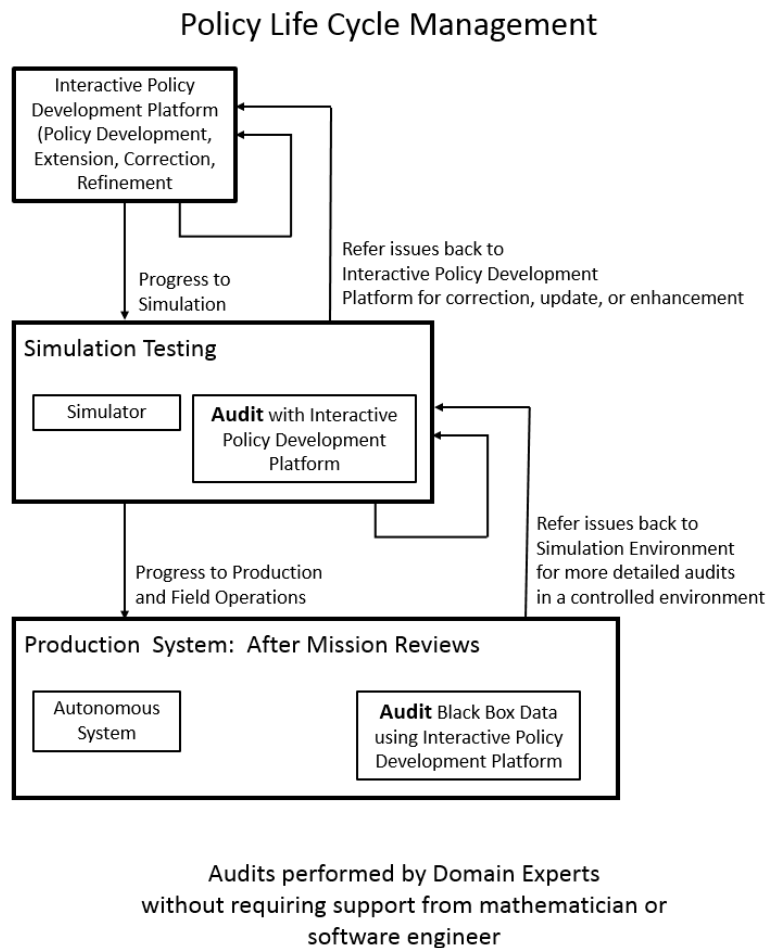


Figure 9.5. After Action Review Audit (Response to Issues and Changes in Adversarial Tactics)

Public Representatives, Lawmakers, Elected Officials Responsible for Using Autonomous Systems

This group is the ultimate customer that is interested in doing more at less cost, since they will be paying the bills. They will want to be assured that the performance of the autonomous system agrees with their view of the world. They will make the ultimate cost-benefit appraisal the way they vote for their own representatives. They will be interested in the public perception of the value of the autonomous system to themselves. This group will probably not have access to a detailed audit, unless things go very wrong, and then only through a detailed public review.

Conclusions

Autonomous systems are here to stay. They will go through growing pains and things *will* go wrong. It is mandatory that auditability and traceability be considered as a primary design goal, because these systems will be addressing more complex problems, and humans will be distanced from more of the real-time decisions. There must be cost-effective ways to audit the behaviour of the autonomous systems, which implies that almost-immediate problem analysis reports need to be generated so that mass-produced devices can be taken offline if systemic problems are identified. When human-provided information causes inappropriate behaviour of the autonomous system, that information needs to be traceable to the source so that new interlocks can be inserted into the operational structure. All of this suggests that auditability and traceability need to be considered as primary deliverables of any autonomous system in which decisions and actions are important.

Finally, auditability and traceability can only be successful in the long term if the tasks are simple and cost effective. If auditing the behaviour of a system is too difficult or time consuming to accomplish, or if it requires too much specialised knowledge to perform, it will probably never be implemented. This will lead to mass-produced misbehaviour.

Recommendations

Senior defence officials should be aware of new technologies that will accelerate the deployment of autonomous systems into the field. This is a real challenge, because the ‘experts’ only know what they know, and what

they don't know is a threat to their expertise. They should also recognise that humans will still be responsible for the behaviour of these systems both (1) for their embedded policies and (2) for information that they provide to the autonomous systems that will contribute to their behaviour.

If and when these systems are armed and assigned engagement goals, the officials need to decide who is responsible for creating the policies that control these systems and assigning values that will control their behaviour. The tendency may be to outsource all software to contractors that supply the hardware.

One recommendation is that autonomous systems architectures be partitioned so that cognitive policies that define the behaviour of the systems can be developed and managed by military personnel. This would also mean that military personnel need to know how to accomplish this task.

One of the most critical activities, whether the systems are armed or not, is the formalisation of a value system. This will vary from one system to another, but there is an apparent reluctance to assign specific numbers to elements such as mission importance, collateral damage, and risk to different parties. While the principle of proportionality may satisfy humans with terms like 'anticipated' military advantage and 'excessive' force, a machine cannot understand these terms. These systems will be constantly balancing risk and reward (pros and cons) to accomplish their goals, and they cannot effectively perform their tasks unless they have a value system that they can understand (based on numbers).

Auditing the behaviour of these systems will be mandatory because things will go wrong that can only be fixed through the auditing process. Auditing failures will not always mean that an autonomous system did not perform correctly. Changes in adversary tactics, weapons, and tools will demand constant evaluation to ensure that the best policies, systems, and weapons are deployed. Autonomous systems should be considered as evolutionary tools under the control of humans; therefore the humans will (hopefully) dictate their evolution.

Because modifying the behaviour of autonomous systems will be a continuous process, the cost of auditing that behaviour needs to be considered as part of the life-cycle cost of these systems. Unlike spending months reviewing the black boxes on commercial aircraft, there will be demands to complete audits in hours and days. So the topic of auditability and traceability should be considered at the beginning of the design cycle rather than as an afterthought.

References

- Chairman of the Joint Chiefs of Staff (CJCS). *Instruction 3160.01A: No-Strike and the Collateral Damage Estimation Methodology*. 12 October 2012. Available at <https://info.publicintelligence.net/CJCS-CollateralDamage.pdf>, accessed 29 March 2015.
- Compsim. 'About KEEL Technology', ND. Available at <http://www.compsim.com/AboutKEEL.htm>, accessed 29 March 2015.
- Compsim. 'Auditing a KEEL-based Policy', ND. Available at <http://s316758578.onlinehome.us/PublicMovies/AuditablePoliciesForAxS/AuditablePoliciesForAxS.html>, accessed 29 March 2015.
- Considine, Glenn D. (ed.). *Van Nostrand's Scientific Encyclopedia*, 6th ed. New York: Van Nostrand Reinhold Company, 1983.
- Department of Defense (DOD). *DOD-STD-2167: Defense System Software Development*, 4 June 1985. Available at <http://www.product-lifecycle-management.com/download/DOD-STD-2167A.pdf>, accessed 29 March 2015.
- Lockheed Martin, 'A Digital Jet for the Modern Battlespace', ND. Available at <https://www.f35.com/about/life-cycle/software>, accessed 29 March 2015.
- Open Web Application Security Project. 'Code Review Metrics', ND. Available at https://www.owasp.org/index.php/Code_Review_Metrics#Inspection_Rate, accessed 29 March 2015.
- NAVSEA. 'Systems Safety Engineering'. *Leading Edge* 7/3. Available at http://www.navsea.navy.mil/nswc/dahlgren/Leading%20Edge/System%20Safety/01_Introduction.pdf, accessed 29 March 2015.
- Salem-Schatz, Susanne, Ordin, Diana, and Mittman, Brian. *Guide to the After Action Review*, 2010. Available at http://www.queri.research.va.gov/ciprs/projects/after_action_review.pdf, accessed 29 March 2015.
- Taleb, Nassim. *The Black Swan: The Impact of the Highly Improbable*. Random House, 2007.

PART 4:

PERSPECTIVES ON IMPLEMENTING AUTONOMY IN SYSTEMS

Four Challenges in Structuring Human-Autonomous Systems Interaction Design Processes

Pertti Saariluoma

Abstract

Human technology interaction provides a wide perspective on technology design. The highest-level designers have four basic design questions to solve. The first question concerns the technical interaction design that defines the functionalities and means of control of a technical artefact. The second question considers how to make technical artefacts easy to learn and use, and how people can best use them (e.g., how to eliminate complexity in using an artefact). The third issue concerns emotional processes in using the technology, for example, how it is possible to gain the trust of users? Finally, designers must ask the ‘*what for*’ question. For what purposes is the technology designed? What are the usages or given technological concepts? This chapter discusses how these questions can be used as a framework to guide individual and organizational design thinking in the context of autonomous systems.

Introduction

Autonomous systems represent the newest emerging breakthrough in information and communications technology. In brief, autonomous systems are technologies with the capacity to perform tasks that previously required human operators to contribute the higher cognitive processes associated with human thinking.⁽¹⁾ Such cognitive processes include categorization, concept formation, learning, judgment and inference, decision making, and problem solving.⁽²⁾ Typical prototypes for autonomous systems that have the capacity to solve complex problems are chess machines, which are able to surpass the

1. Newell and Simon 1972; Veres and Molnar 2010.

2. Kahneman 2011; Newell and Simon 1972.

performance of even the best chess players. Thus autonomous capacities enable technical artefacts⁽³⁾ to respond in a rational manner in unforeseen situations. Industry is awakening to the promises of autonomous technologies. Google has its self-driving cars, and Facebook plans to connect all of its drones into a huge network.⁽⁴⁾

However, not all automatic technical artefacts represent an autonomous system. Traditional stimulus/response-type technical artefacts are not autonomous systems. For example, a door that opens when it registers human body temperature or movement is an automatic system and independent from users' continuous control, but it is not an autonomous system. If it had the capacity to decide for which people it should open, for instance depending on the gravity of their illness (e.g., Alzheimer's disease) or personality type, it could then be described as having autonomous capacity as it would be conducting a demanding (autonomous) categorization task. Thus while there is an overlap between automatic and autonomous systems, in practical contexts the two types are different enough to be considered separately.

Autonomous systems can replace people in many complex tasks, for example to drive cars; to operate weed control machines, tractors, or drones; to control air traffic or factory lines; to run offices; and to take care of elderly people and make life safer for people suffering from many kinds of illnesses. Indeed, their possible applications could improve the lives of numerous people.

However, even autonomous systems need people – all machines and technical artefacts do. Autonomous systems may be more complex and more independent of immediate human control, yet people are needed to design, program, build, parameterize, and give the systems rational goals, for example. But human–technology interaction (HTI) problems are inevitable. By carefully considering the design of such interactions, it may be possible to realise the human dimension of autonomous technologies without unnecessarily risking human lives. As human control of autonomous technologies becomes swiftly more complex, operations become very difficult to perform. For example,

3. Generally speaking, a 'technical artefact' is any human-made tool, machine, device, instrument, programme, or natural process that people use to facilitate their actions. See <http://plato.stanford.edu/entries/artefact/> for further details.

4. Debord 2014.

it is easy for a juggler to play with one ball, even with two or three, but it is impossible to juggle with 50 balls.

To get the best out of the new technologies, it is essential to critically analyse and reconsider the ways and paradigms of previous interaction thinking. New technologies will most likely be guided by new kinds of interaction design thinking, but they must be built on what is already known about HTI design. The old paradigms define the questions that must be solved by the designers of autonomous systems, but the thoughts must be organised in a new way.

Scientists often use various frameworks of thinking and assumptions, called paradigms.⁽⁵⁾ Researchers working within a paradigm share questions, concepts, methods, and theories. The main paradigms in HTI research have been ergonomics, human computer interaction (HCI), and user experience (UX). Today, the field of HTI appears quite confusing, as it is full of partly overlapping approaches and paradigms such as usability, HCI, Kansei-engineering, UX, entertainment computing, or designing for pleasure. To meet the challenges of designing autonomous technologies, it is essential to consider the main paradigms of HTI design in a systematic and critical manner, to find them a place in holistic design thinking.

This chapter first reviews the idea of a ‘paradigm’ in science, and reflects on the need to ask whether current paradigms are sufficient for examining autonomous systems. Second, four fundamental questions are proposed as a lens through which to examine the issues related to HTI and autonomous systems. Finally, it concludes with a description of an ideal for design thinking.

History of HTI Design Thinking

Scientific studies on HTI started in the mid-1800s,⁽⁶⁾ and since then the interest in different aspects of HTI has gradually increased. During this time, scientists have constantly encountered new theoretical and practical problems; solving these problems has generated new paradigms: human factors, physical and cognitive ergonomics, emotional usability, and activity theory.⁽⁷⁾ Given that they were often developed alongside each other, these paradigms often

5. Kuhn (1962) defines paradigms as ‘universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of researchers’.

6. Karwowski 2006; Moray 2008.

7. Card, Moran, and Newell 1983. Karwowski 2006; Norman 2002.

overlap; successive paradigms can form a research programme with a common challenge and discourse.⁽⁸⁾ Thus researchers in ergonomics may be interested in the emotional aspects of usability, and HCI practitioners may apply ideas of cognitive ergonomics in their work. The researchers commit themselves to certain approaches and criticise their competitors, but often recognise that it is rational to seek answers outside their own paradigms. Although these competing paradigms overlap, they use different concepts and lack a common standard of measurement.⁽⁹⁾ Kansei engineering and user experience research or human factors and ergonomics are typical examples.⁽¹⁰⁾

The main paradigms of HTI can be understood based on what questions they originally raised. It is essential to see what kinds of concepts (and, consequently, what kinds of questions) have been important in the different paradigms. Since each of the major paradigms differs in how it perceives the most fundamental questions of the field, the main paradigms of HTI research are mainly complementary. Each of them adds something essential to the problems of the field, and thus offers a new perspective.

Analysing basic questions is also important because it reveals whether there are some important problems related to HTI that have not yet been addressed by the researchers and designers. Indeed, it is always essential to ask whether the given intuitive foundations really address all the questions. Scientific progress is characterised by the critical analysis of the foundations of existing paradigms.⁽¹¹⁾

This kind of review leads to a series of questions that are necessary to understand the field as a whole and to meet the design challenges posed by new technologies such as autonomous systems: Do the main research paradigms address all the essential issues in HTI, or are new ways of considering those basic problems needed? Is more clarity on the foundations of HTI thinking needed, and if so, why? Can current ways of thinking address all the aspects of HTI that are essential in design?

This investigation is pursued to illustrate that there are major questions underlying all areas of HTI research. They have emerged at different points in

8. Lakatos 1968.

9. Feyerabend 1975; Kuhn 1962.

10. Hassenzahl and Tractinsky 2006; Nagamashi 2011.

11. Saariluoma 1997.

time in the discourse to aid practical design, and today these questions define the field itself by highlighting the challenges and problems that designers must solve.

When working with any kind of HTI process, designers must always meet the challenges surrounding four foundational questions.⁽¹²⁾ First, they have to be able to define the technical interaction instruments that users need in order to control a particular technology. Second, they have to find ways to enable people to easily use machines and other technical artefacts. Third, they have to solve emotional problems typical to a particular technology. Finally, they have to be able to find out how people use the technology in their lives.

Challenge 1: Technical User Interaction

User guides for technical artefacts typically contain information about what these tools can achieve, and how to do so. Mobile phones, for example, may have global positioning systems (GPS) maps, GPS data, and the ways people can use these applications. User instructions detail how to navigate through menu systems or command paths such as ‘GPS/maps/navigator’; touchscreen technologies have made these paths shorter. These instructions tell users how they can get the technical artefacts in the state they expect them to be, for example, a state in which the GPS guides their car’s movements in an unfamiliar city. Underneath the series of technically defined user operations and commands, one can find a series of states and transitions. These state-transition series form event trees, which define how people can get technical artefacts into the expected state.

The purpose of technology is to allow a user to exploit a given technology to realise his or her action goals. The role of technology in human action is called ‘functionality’, which has become central, as ICT devices can have myriad potential functionalities. Functionalities define how people can get a technical artefact to work in an expected manner, and event trees stipulate how the goals of different functionalities can be reached, i.e., what states in the interaction tree will transform the technical artefact from the initial state to the goal state. Especially important here are the states that require active human involvement in the transformation process.

12. Saariluoma and Leikas 2012.

All technical devices have an event-tree structure. Computers are navigated from state to state, and although cars have ‘analogue’ steering wheels, they can be seen as ‘digitalised’ state transition systems. Autonomous technologies are new in the sense that they have the capacity to operate in looser contact with human controllers than earlier technical artefacts. Nevertheless, human involvement is unavoidable, as people must define the goal states of the systems. While their operation may be more complex to control (because much of the ‘intellectual’ work is transformed from people to machines),⁽¹³⁾ technical interaction design is ultimately based on functionalities and event trees.

The tasks and goals of autonomous systems may be relatively openly designed so that the systems can themselves decide what to do. A robot searching for an amnesiac patient in a large forest must be able to discriminate wolfs and bears from the lost human. Thus it must have the capacity for the rather complex semantic analysis of sensor information. Many simple input signals may vary: the body temperature of the lost person may be altered, or she or he may have lost consciousness. The system must also be able to carry out its task if contextual parameters vary, from frost to a forest fire, for example.

In autonomous systems (or indeed any use of technical artefacts), users need to have the technical means (i.e., the controls) to communicate with the artefacts.⁽¹⁴⁾ The controls are used to give information to the technical artefacts about the expected goal states. Essential to control is to offer the possibility for machines to choose between alternative courses of action to reach the given goals; while machines cannot set their own goals, they can selectively pursue them once assigned.

In addition to these controls, event paths, and instructions that define how to reach the expected artefact states, it is essential to define feedback information for the users during the design process.⁽¹⁵⁾ This calls attention to such problems as what languages a navigation system should use, whether it should use a male or female voice, how it describes the location (i.e., does it use bird’s-eye perspective), and what kind of additional information it should give to users about locations. The semiotics of the interface are also essential

13. Slaughter, Giles, and Downey 2010; Veres and Molnar 2010.

14. Cooper, Reimann, and Cronin 2007; Pahl et al. 2007.

15. Cooper, Reimann, and Cronin 2007.

here.⁽¹⁶⁾ Feedback information can be directly connected to the use of the technology, or it can simply inform the navigation process about the state of the technical artefact so that the user can go toward the expected goal state.

The first question in HTI research thus concerns how to use a technology – or, more precisely, what kinds of functionalities technologies offer to users. This is the fundamental question of technical usability. Answering it requires designers to define the functionalities of a technology and its interaction components. In essence, the question is to define how (i.e., using what kinds of interactions, operations, or events) people can get a technical artefact to reach the expected goal state. For example, a mouse click on a search button may open a set of relevant windows or pressing a brake button may stop a crane. With respect to autonomous systems, the ‘span’ between presenting a command and eventually reaching the goal is much wider and more open than in the case of technical artefacts that are continuously controlled.

The technical design of autonomous systems should give people control over the goals of their actions. They have to be aware of the states of the system, and to have a veto over all life-critical operations and changes in targets. In such systems, technical user interfaces represent a hard task for interaction designers, as it is essential to organise them on a higher conceptual level than for continuously controlled systems. People do not constantly control an autonomous system; they only get involved in critical issues. This is why the technical interface must provide new ways to get information about the critical issues and the systems’ anticipated behaviours. Crucial design questions include: what is the task, what are the obstacles, and should the goals be changed. Such high-level operational issues must be defined in advance in order to design the functionalities and information to increase the effectiveness of the interface elements in a rational manner.

Challenge 2: Are People Able To Use the Autonomous Systems?

Technologies are of little use unless people are able to employ them. In practice, a technology may be so complex (and thus too difficult to control) that it causes users to make errors. Or using the technology may take too much

16. De Souza 2005.

time, cost too much, or involve certain dangers and risks if used inappropriately. For these kinds of reasons, the problems of ‘can use’ or ‘being able to use’ a technology have become important in interaction design thinking.⁽¹⁷⁾ The usability of autonomous systems is a challenge of primary importance to designers.

Several scientific approaches have sought to understand the components of good usability. Typical examples have been human factors, ergonomics, HCI,⁽¹⁸⁾ and, more recently, inclusive design and usability engineering.⁽¹⁹⁾ In these paradigms, researchers have incorporated knowledge, concepts, and methods from various basic sciences such as industrial engineering, medicine, psychology, organisation research, and communication research to solve the problems of ‘can use’.

There are subtle differences between these paradigms. HCI, for example, focuses much more intensively on the problems of human and computer interaction compared to ergonomics or human factors. Given the importance of computers and computing in modern technology, however, interacting with computers is directly or indirectly important to all paradigms. Consequently, the paradigms differ little in their interest in the question of how technologies can be used. For the sake of simplicity, the term ‘ergonomics’ can be used to refer to all the traditions concentrating on ‘can-use’ problems.

Many of the problems can be researched in the context of psychology. Psychology-based interaction research can generally be called ‘user psychology’, and the human dimension of these problems can often be addressed using psychology-based studies.⁽²⁰⁾ The core idea is to redefine a given design problem, such as complexity, in terms of psychology and then to analyse it using concepts of modern psychology such as the theory of the limited capacity of attention and working memory.⁽²¹⁾

Ergonomics is commonly divided into physical, cognitive, cultural, and organisational domains.⁽²²⁾ Physical ergonomics studies human anatomical,

17. Card, Moran and Newell 1993; Karwowski 2006; Nielsen 1993; Norman 2002.

18. Bridger 2009, Karwowski 2006; Moray 2008.

19. Nielsen 1993.

20. Moran 1981, Saariluoma 2004, Saariluoma and Oulasvirta 2010.

21. Baddeley 2007, Broadbent 1958; Covan 2000; Miller 1956.

22. Karwowski 2006.

physiological, biomechanical, and anthropometrical issues, which are directly related to physical activity. Thus physical problems of using technologies are solved using anatomical, psychological, and physiological knowledge. Often this is assumed to be irrelevant in designing autonomous systems. However, in practice there are situations in which technical human control is needed to set goals and define the parameters of using the systems.

Human control of technical artefacts is, at the lowest levels, based on perceptual motor cycles and eye-limb co-ordination.⁽²³⁾ This is the required level for designing human autonomous systems interaction (HASI) processes. For example, a ‘bewildered’ system must be stopped very fast, and the speed of human response depends on the ergonomics of the controls. Thus the higher levels of human cognition (i.e., attention, memory, language, and thought) are central in designing human control of autonomous systems, as these systems can be very complex from a human point of view.

The role of higher cognitive processes has most often been analysed according to the paradigms of cognitive ergonomics or cognitive skills research.⁽²⁴⁾ Typical examples of higher-level problems are mental workload, communication, decision-making, skilled performance, and situation awareness. One can also include cognitive ergonomics and the rising field of neuroergonomics on the list of relevant research approaches.⁽²⁵⁾

The application of knowledge about the limited capacity of human attention is one example of this approach. People can normally attend only one target at a time⁽²⁶⁾ and remember a few new items.⁽²⁷⁾ Limited capacity or mental workload is significant in HTI, as it significantly slows down the speed of information search and selection, and limits what humans can remember about the task requirements. Thus since the performance on any task may be disturbed by the limited capacity to process information, interaction solutions must support circumventing the limits.

Interactions are ‘dialogues’ between the technical artefact – machine, device, or programme – and the human user. Dialogues are uniquely defined

23. Bridger 2009; Schmidt and Lee 1988.

24. Nielsen 1993; Norman 2002.

25. Parasuraman and Rizzo 2008.

26. Shiffrin and Schneider 1984.

27. Broadbent 1958; Covan 2000; Miller 1956.

to consist of a limited set of signs. Each sign is associated with an event, i.e., a process leading to a definite goal state. The human role is to make the decisions between different goal states, to describe them to the system, and to steer the artefact to the desired state. Sometimes dialogues are simple: turning a power switch on activates the electric current in a machine and makes it possible to use it. Dialogues can also be multi-dimensional, such as when controlling an autonomous submarine.

To make the dialogues possible, it is essential to design interaction elements for a user interface. They can be input elements launching event flows, or output elements indicating the states of different processes. In design language, they can be called controls, meters, widgets, icons, or text boxes, for example. Their function is to let users give commands to the artefacts and to inform them about the state of the artefact. Thus interaction elements are signs in dialogues between machines and people.⁽²⁸⁾

Another important element is situational awareness, which refers to a user's capability to be knowledgeable of the relevant features of the action environment.⁽²⁹⁾ This means that people are able to both detect the necessary perceptual cues and understand the meaning of these cues for their action. A drone controller must be aware of the state of the plane, its spatial positions, and the locations of other drones. If they are not fully aware of the situation, there is always the danger of an accident.

Situational awareness is intimately linked with semiotics. In autonomous systems, users are removed from (rather than inside) the artefacts. All of the knowledge about the situation must therefore be collected from data streams, which are strings of semiotic symbols. Thus designing technical semiotics that are easy for humans to understand is the key to creating high-level situational awareness and effective control systems. It is not sufficient to create a good system of codes; the psychological problems of using the semiotic code must also be solved. This is why designing semiotics for interfaces in autonomous systems must be based on multidisciplinary thinking. The field of technological psychosemiotics generates a system of relevant signs based on the analysis of users' mental capacities.⁽³⁰⁾

28. de Souza 2005.

29. Endsley 2011.

30. Saariluoma and Rousi forthcoming.

Finally, the interaction processes that require decision making and problem solving must be investigated. These higher levels of human thinking can often be considered a problem of resilience (i.e., the ability to prepare for, and the capacity to cope with, disturbance and unexpected situations). This notion is very important in designing autonomous systems, as their operations are often difficult to predict. Critical-use situations will easily become typical to interacting with autonomous systems. To deal with the unexpected, people need the capacity to evaluate the consequences of two or more alternatives quickly and accurately (e.g., whether to redirect or destroy a system).⁽³¹⁾ However, situations are not always clear-cut enough that it is possible to take a decision. One needs to understand the alternatives, but if no clear alternatives exist (or if one can see the goal, but does not know how to reach it), it will be essential to clarify the situation. These kinds of situations are typical to human problem solving, and can be addressed by creating suitable support systems during the interaction design process.⁽³²⁾

From a cognitive point of view, HASI design involves defining and solving problems that must be supported by different levels of psychological and human research knowledge. First, designers must rely on perceptual motor cycles.⁽³³⁾ In order to solve the problems of resilience, they also have to work with an increasing workload, consider communication and interaction semiotics, and improve the decision-making and problem-solving processes.

Since autonomous systems represent highly unpredictable technical artefacts, designing their usability will require new thinking in many respects. Here, the focus has been on defining the main psychological problem levels. Though the presented levels of design problems have been considered one at a time (and thus they look very equal), higher cognitive processes such as decision making and thinking will be the main challenge when designing the usability of autonomous systems.

31. Kahnemann 2011.

32. Newell and Simon 1972.

33. Schmidt and Lee 1988.

Challenge 3: Do Users Like Autonomous Technologies?

The next challenge of HTI design was when art designers, marketing researchers, and others began to think that products must be pleasant to use and desirable – a goal termed ‘like to use’. They must be beautiful, elegant, and reflect the personality and social position of their users⁽³⁴⁾ in order to create great user experiences.⁽³⁵⁾

‘Liking’ can also be seen as arousing (among users) positive emotions and a willingness to use and buy products. Although this emotional connection is vital for marketing the products, the emotional elements of interaction design are still relatively under-researched. The four best-known traditions are: (1) Kansei engineering,⁽³⁶⁾ (2) emotional usability, (3) fashionable UX, and (4) entertainment computing research.⁽³⁷⁾ First, Kansei engineering is a design system used to give products affective attributes – for example, the notion that sports cars should be powerful, youthful, energetic, and elegant. The challenge is incorporating these elements into cars and their parts; the task of Kansei engineering is crucial to find out which design parameters give users such impressions.

Second, the emotional usability approach⁽³⁸⁾ is different, since emotional design can be attached to explanatory physiological, psychological, sociological, and cultural theories.⁽³⁹⁾ Thus it involves not only asking *what kinds* of designs are more beautiful or distrusted, but also *why*.

The third important tradition is UX research. The ability to design for user experience became a focal issue in HTI design in the early 2000s.⁽⁴⁰⁾ UX describes a user’s overall experience when employing a product or system.⁽⁴¹⁾ It includes users’ perceptions of HTI, which, in turn, are associated with several inter-related factors ranging from traditional usability to aesthetic, hedonic, and affective aspects of technology use. There is no specifically common definition of UX; the notion is mostly used to extend the viewpoints of usability

34. Nagamashi 2011.

35. Hassenzahl and Tratinsky 2006; Norman 2004.

36. Nagamashi 2011; Rauterberg 2006, 2007.

37. Hassenzahl and Tracktinsky 2006; Norman 2004, Rauterberg 2006.

38. Nagamashi 2011; Norman 2004.

39. Saariluoma 2004, Saariluoma and Oulasvirta 2010.

40. Hassenzahl and Tracktinsky 2006.

41. Cooper, Reimann, and Cronin 2007; Kuniavsky 2003.

and human-centred design. UX research is based on the understanding that the experience of even simple artefacts does not exist in a vacuum, but is realised in a dynamic relationship among people, environments, and objects. The concept of UX is thus broadened from traditional usability to involve different aspects of experiencing products or services, including physical, sensitive, cognitive, and emotional relations.⁽⁴²⁾ Hassenzahl and Tractinsky explain UX as a consequence of three categories of factors: those related to a user's internal state, the characteristics of the designed system, and the context within which the interaction occurs.

UX as a research paradigm shifts the emphasis from the functional features of HTI to the quality of interaction. Norman stresses the affective and emotional aspects of interaction, while McCarthy and Wright emphasise the meaning of culture in experience.⁽⁴³⁾ This view brings a social aspect in terms of co-experience and shared experiences to the concept of UX. The emphasis of the approaches to UX mentioned above is on people's experiences, feelings, (positive) emotions, perceptions, and behaviours during use.⁽⁴⁴⁾ These attempts to delineate user experience have been based either on cognitive science and task-related experimental analysis, or on phenomenological approaches and qualitative analysis.

UX design (or interaction design) thus aims to measure and enrich the user experience (i.e., the product quality)⁽⁴⁵⁾ by focusing on the interactions between people and products or services, and the experience resulting from this interaction. To create positive HTI, it focuses on 'how to create outstanding quality experiences rather than merely preventing usability problems.'⁽⁴⁶⁾ This is a natural trend in the design that responds to the need to deal with all kinds of interaction technologies in addition to computer technology. It reflects the broadening of the focus from computers and work-related tasks to the usage of a wide range of interactive technologies, bringing the user's overall and lived experience and natural interaction into the focus of the design.⁽⁴⁷⁾ An essential

42. Kuniavsky 2003.

43. Norman 2004; McCarthy and Wright 2004.

44. Hassenzahl and Tractinsky 2006.

45. McCarthy and Wright 2004.

46. Hassenzahl and Tractinsky 2006, 95.

47. McCarthy and Wright 2004.

challenge for the design is to understand the factors that constitute and shape users' experience of new technologies and smart environments.

UX addresses the factors that influence the usage of products and services, such as user satisfaction and acceptance.⁽⁴⁸⁾ Negative HTI factors include the rejection, total abandonment, or partial non-acceptance of technology. Different technology acceptance models include elements of user attitudes toward (and experiences of) technology, as well as elements of different personal and environmental characteristics. The first models of user acceptance focused on the interaction between the user and the technology. The later models also consider the change in people's acceptance of technology along with (and due to) the usage of products and services.

Other important factors in addition to user satisfaction and acceptance are the system's intuitiveness, transparency, enjoyability (e.g., the product's emotional, hedonic and practical benefits), and the user's sense that he or she is in control. Jordan argues that instead of stressing usability in the design, the focus should be on making products a joy to own and use⁽⁴⁹⁾ – i.e., a pleasurable interaction between a product and a person.

Although products have been designed according to HCI models and tested for high usability, they do not necessarily become closely linked with people's lives on an emotional level. Emotions form an essential part of the user experience, and can be even more important than usability for a product's success, as emotions create satisfaction and an awareness of the product or brand.⁽⁵⁰⁾

Emotions are closely related to a number of other concepts such as feelings and affects. 'Feeling' is often used in colloquial language to describe a broader concept than emotions.⁽⁵¹⁾ Affect, also closely linked to emotion, has mainly been used in the older psychological literature. Emotions are also closely related to needs and motives:⁽⁵²⁾ people value things and make decisions about their behaviour based on their emotions. Therefore, emotional design is an essential element of natural interaction design.

48. Venkatesh 2000.

49. Jordan 2000.

50. Hassenzahl and Tractinsky 2006.

51. Oatley, Keltner, and Jenkins 2006.

52. Norman 2004; Rauterberg 2006; Saariluoma 2004.

Another valuable quality attribute of design is aesthetics. The research into the relationship between aesthetics and HTI is still in its infancy, but it has produced some important finding. For example, consumers are sometimes more interested in aesthetic pleasure than in usability,⁽⁵³⁾ and what the product says about its owner is an important factor in purchase decisions. Aesthetic value is heavily dependent on trends and fashion, and differs among cultures and different lifestyles.

The paradigms discussed above demonstrate the importance of ‘liking to use’. Liking is an emotional issue that highlights the importance of human emotions in interaction-design thinking.⁽⁵⁴⁾ Autonomous technologies create many promises, such as freeing people from many kinds of simple tasks by replacing their work with autonomous systems. Yet it is important to consider the perspective of the whole society regarding a branch of technology or a type of technical device. Even in civil tasks one can ask whether people trust the technology at hand. For example, do passengers like to use automatic metros? Do citizens like autonomic traffic? What happens to trust if systems fail? What are the social consequences? What if technologies such as ‘killer drones’ take control? Designers must thus take these socio-emotional aspects of autonomous systems into account.

The first example of societal feelings toward a type of technology can be found in nuclear technology. People lost their trust in nuclear power stations after accidents in Three Mile Island (1979), Chernobyl (1986), Forsmark (2006), and Fukushima (2011).⁽⁵⁵⁾ Consequently, many governments had to change their nuclear energy policies.

When military users operate autonomous systems on a battlefield thousands of kilometres away, the true consequences of the actions may be easily blurred. Therefore, it is important to have a clear feedback system. The operative user must keep in mind the emotional aspects of the situation in order to assess the rationality of the actions: war-time killing must not appear to be ‘unreal’, like a computer game for example. If the emotional link to the consequences of the actions is lost, events can spiral out of control, and lead to an outcome that jeopardises the goals of the entire operation.

53. Hassenzahl and Tracktinsky 2006.

54. Khalid and Helander 2006.

55. Visschers and Siegrist 2012.

Challenge 4: Life-Based Design

Technology's only role in modern society and human life is to enhance people's ability to reach their goals. Thus technology design, and HTI design in particular, can be described as designing life or designing *for* life – or even *of* life. HTI design that begins by analysing life can be called life-based design (LBD).⁽⁵⁶⁾ The goal of interaction design goes beyond the immediate capacity to use technology; it involves figuring out how best to incorporate technical devices and services into human life actions.

Traditionally, technology has been seen as value neutral, associated only with facts, and outside the realm of ethics. Yet more and more, designing technology is taking into account the ethical dimensions of each device.⁽⁵⁷⁾ Von Wright, for example, deliberates on the ethics of technology in respect to real humanism, human beings, and sociality.⁽⁵⁸⁾ These discourses have involved perspectives such as value-oriented design, value-sensitive design, worth-based design,⁽⁵⁹⁾ ethical design, and technology ethics.⁽⁶⁰⁾ In all these approaches, the attitude and the will of the designer toward the design activity have a much greater effect on the outcome of the design than merely mechanically complying with certain design guidelines and rules. None of these approaches brings a holistic approach to designing to life, as they are limited to what the goals of the actions *should* be rather than what they *are*.

LBD is a multi-dimensional and holistic approach that emphasises the importance of understanding people's lives as a basis for creating designs. It involves planning for the social consequences of autonomous technologies. For example, if an automated car is expected to replace millions of taxi, bus, and lorry drivers, there should be plans in place to help them apply their skills in new areas; the same would apply to people in related fields, such as mechanics.

LBD's core aim is to release technology that will be widely accepted by people. It does so by exploring people's forms of life, values, and everyday contingencies such as age, family and marital status, social status, profession, health issues, education, gender, and skills. These factors ultimately impact everyday needs,

56. Leikas 2009; Saariluoma and Leikas 2010.

57. Stahl 2006.

58. Von Wright [1981] 2007.

59. Cockton 2004.

60. Boven 2009.

for example those related to communication and companionship. The goal is to analyse the role and function of the technology in life. These factors guide the HTI design processes with the help of technology-supported actions.⁽⁶¹⁾ The design issue is to find out how the quality of people's lives can be improved. It takes an holistic approach, which means that all design issues in LBD are biologically, psychologically, and socio-culturally motivated.

The form-of-life analysis should generate *human requirements* for the product or service, which define how people's life should be improved in a specific way. The human requirements create the basis for the next phase in the design by introducing the design theme and the explanatory facts behind it.

Autonomous systems do not necessarily create new goals for living, but they provide new means to reach them. The analysis of 'what for' and 'why' must lead to concrete analysis of how to improve individuals' lives. Thus, the focus of thinking is on designing new or improved forms of life than on the actual technical artefacts. The function of technical artefacts is to make new ways of acting possible, and thus allow people to reach their goals.

Interaction Design Thinking for Autonomous Systems

Much of the research in HTI research can be seen in the context of four relatively independent challenges. They are all defined by their basic questions. The first challenge is technical users' interface, or 'how to get technology to do what it should do'. The second question concerns the obstacles to using technologies: 'can we use or are we able to use' technologies. 'Can' in this broad sense refers to insufficient or inadequate skills to use as well as to any obstacles to their smooth and efficient use. The third question refers to emotions in using technologies. This is the issue of 'how we like' to use technologies. The final challenge is the 'what for' – understanding for what purposes a technology can be used.

Since autonomous systems have the capacity to perform tasks that previously required active human participation, designing the human dimensions for these kinds of technologies is tricky. The four basic discourses or challenges discussed above enable designers to think holistically and systematically about how to carry out human autonomous system design processes.

61. Leikas 2009; Saariluoma and Leikas 2010, 2012.

Since designing and innovating are in fact human thought processes, to enable people to design in a rational manner it is necessary to think how to control such thought processes.⁽⁶²⁾ Thus designers should ask: how to make it easy to input parameters into a system, how to make goal setting fast, or how to make systems learn from one another.

Here, we are one step from the crucial insight. At a higher level, most of the technical artefacts have the same conceptual structure.⁽⁶³⁾ Designers must formulate, ask, and answer the most basic questions in order to generate (and answer) the product-specific design questions. Thus, design processes can be structured by an ontology of questions and answers,⁽⁶⁴⁾ such as those discussed above.

Human autonomous systems interaction can be structured using ontological thinking. Ontologies provide to guide thinking, for example helping people use earlier experiences and domain-specific concepts to create new designs.⁽⁶⁵⁾ Ontologies are needed to express basic concepts of some area to build relevant knowledge systems or to provide support for designers' thinking processes and to transmit knowledge from one design process to another. Content ontologies may help designers by explicating parts of the domain-relevant knowledge; thus previously tacit aspects of the domain can be subsumed into explicit social discourse and critical analyses. Ontologies may also help people remember all the relevant aspects of a design process and communicate this information to other designers.

Many companies make continuously new versions of their old products, for example consumer products such as cars, computers, or mobile phones. Rather than starting from scratch for each new version, they use old plans to develop new products and change only the essential design points. Conceptual ontologies can help convey design knowledge from one generation of products to the next.⁽⁶⁶⁾

To find out what kind of ontology could be used in HTI design, it is essential to discuss the contents of the main categories, and in what kind of knowledge

62. Cross 2001; Simon 1969.

63. Minsky 1967.

64. Saariluoma and Leikas 2010, 2012.

65. Gero 1990.

66. Saariluoma and Leikas 2010, 2012.

management tool they could be implemented. Here, it is important that design process can be seen as a structure of questions that can activate goals and thought processes. While such questions allow the use of previous solutions, they also trigger a search for new answers and new questions.⁽⁶⁷⁾ Question-based ontological design thinking thus helps designers manage their actions in an organised manner.

The four main design problems discussed above form the basic ontological structure of any human interaction design process. The designers have to: define what the technical artefact is supposed to do, find the most effective and easiest way of doing it, solve the three relevant emotional problems (e.g., trust), and incorporate the new technologies into human life. The task of human-autonomous systems research is to explicate the underlying design-question structures.

Thus, the analysis has led to a system that can be used to direct and manage innovative design thinking in considering HTI processes when creating autonomous systems. The system applies cognitive science theory and conceptual and metascientific analysis while relying on modern psychology. This approach defines a holistic system for thinking about what is perhaps the most complicated emerging technology in history.

67. Ibid.

References

- Baddeley, A. *Working Memory, Thought, and Action*. Oxford: Oxford University Press, 2007.
- Boven, W. R. *Engineering Ethics*. London: Springer, 2009.
- Bridger, R. S. *Introduction to Ergonomics*. Boca Raton, FL: CRC Press, 2009.
- Broadbent, D. E. *Perception and Communication*. Elmsford, NY: Pergamon Press, 1958.
- Card, S. K., Moran, T. P., and Newell, A. *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Erlbaum, 1983.
- Chandrasekaran, B. Josephson, J. R., and Benjamins, V. R., 'What are Ontologies, and Why Do We Need Them?' *Intelligent Systems and their Applications* 14 (1999): 20–6.
- Cockton, G. *Value-centred HCI*. Paper presented at the Third Nordic Conference on Human-Computer Interaction, Tampere, 26–27 October 2004, 149–60.
- Cooper, A., Reimann, R., and Cronin, D., *Modeling Users: Personas and Goals. About Face*. Indianapolis: Wiley, 2007.
- Covan, N. 'The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity'. *Behavioural and Brain Sciences* 24 (2000): 87–185.
- Cross, N. 'Designerly Ways of Knowing: Design Discipline Versus Design Science'. *Design Issue* 17 (2001): 49–55.
- Debord, M. 'You Won't Even Know These Facebook Drones Are in the Sky'. *Business Insider*, 19 September 2014. Available at <http://www.businessinsider.com/you-wont-even-know-these-facebook-drones-are-in-the-sky-2014-9>, accessed 23 March 2014.
- De Souza, C. S. *The Semiotic Engineering of Human-Computer Interaction*. Boston: MIT Press, 2005.
- Feyerabend, P. *Against Method*. London: Verso, 1975.
- Gero, J. 'Design Prototypes – A Knowledge Representation Schema for Design'. *AI Magazine* 11 (1990): 26–36.

- Gruber, T. R. 'Toward Principles for the Design of Ontologies Used for Knowledge Sharing'. *International Journal of Human Computer Studies* 43 (1995): 907–28.
- Hassenzahl, M., and Tractinsky, N. 'User Experience – A Research Agenda'. *Behaviour & Information Technology* 25 (2006): 91–7.
- Jordan, P. *Designing Pleasurable Products: An Introduction to the New Human Factors*. London: Taylor and Francis, 2000.
- Kahneman, D. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.
- Karwowski, W. *The Discipline of Ergonomics and Human Factors*. Hoboken, NJ: John Wiley & Sons, 2006.
- Khalid, H. M., and Helander, M. G. 'Customer Emotional Needs in Product Design'. *Concurrent Engineering-Research and Applications* 14 (2006): 197–206.
- Kuhn, T. *The Structure of Scientific Revolutions*. University of Chicago Press. Chicago, 1962.
- Kuniavsky, M. *Observing the User Experience: A Practitioner's Guide to User Research*. Los Angeles: Morgan Kaufmann, 2003.
- Lakatos, I. 'Criticism and the Methodology of Scientific Research Programmes'. *Proceedings of the Aristotelian Society* 69 (1968) 149–86.
- Leikas, J. *Life-based Design – A Holistic Approach to Designing Human-Technology Interaction*. Helsinki: Edita Prima, 2009.
- McCarthy, J. and Wright, P. 'Technology as Experience'. *Interactions* 11 (2004): 42–3.
- Miller, G. A. 'The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information'. *Psychological Review* 63 (1956): 81–97.
- Moran, T. P. 'Guest Editor's Introduction: An Applied Psychology of the User'. *ACM Computing Surveys (CSUR)* 13 (1981): 1–11.
- Moray, N. 'The Good, the Bad, and the Future: On the Archaeology of Ergonomics'. *Human Factors* 50 (2008): 411–17.
- Nagamashi, M. 'Kansei/Affective Engineering and History of Kansei/ Affective Engineering in the World'. In *Kansei/Affective Engineering*, edited by M. Nagamashi, 1–12. Boca Raton, FL: CRC-Press, 2011.
- Newell, A., and Simon, H. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972.

- Nielsen, J. *Usability Engineering*. San Diego, CA: Academic Press, 1993.
- Norman, D. A. *The Design of Everyday Things*. New York: Basic Books, 2002.
- Norman, D. A. *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books, 2004.
- Oatley, K., Keltner, D., and Jenkins, J. M. *Understanding Emotions*. Oxford: Blackwell Publishing, 2006.
- Pahl, G., Beitz, W., Feldhusen, J., and Grote, K. *Engineering Design*. London: Springer, 2007.
- Parasuraman, R., and Rizzo, M. *Neuroergonomics: The Brain at Work*. Oxford: Oxford University Press, 2008.
- Rauterberg, M. 'From Personal to Cultural Computing: How To Assess a Cultural Experience'. In *uDayIV – Information nutzbar Machen*, edited by G. Kemper and P. von Hellberg, 13–21. Eichengrund: Pabst Science Publication, 2006.
- Rauterberg, M. 'Ambient Culture: A Possible Future for Entertainment Computing'. *Interactive TV: A Shared Experience-Adjunct Proceedings of EuroITV* (2007): 37–9.
- Saariluoma, P. *Foundational Analysis: Presuppositions in Experimental Psychology*. London: Routledge, 1997.
- Saariluoma, P. 'Explanatory Frameworks for Interaction Design'. In *Future Interaction Design*, edited by A. Pirhonen, H. Isomäki, C. Roast, and P. Saariluoma, 67–83. London: Springer, 2004.
- Saariluoma, P., and Leikas, J. 'Life-based Design – An Approach to Design for Life'. *Global Journal of Management and Business Research* 10 (2010): 17–23.
- Saariluoma, P., and Leikas, J. 'Managing Life-Based Design as a Micro-Innovation Process'. Paper presented at the Design Creativity Workshop, 6 June 2012, Texas A&M University, College Station, Texas.
- Saariluoma, P., and Oulasvirta, A. 'User Psychology: Re-assessing the Boundaries of a Discipline'. *Psychology* 1 (2010): 317–28.
- Saariluoma, P., and Rousi, R. 'Symbolic Interactions: Towards a Cognitive Scientific Theory of Meaning in Human Technology Interaction'. *Journal of Advances in Humanities* (forthcoming)

- Schmidt, R. A., and Lee, T. D. *Motor Learning and Control*. Champaign, IL: Human Kinetics, 1988.
- Smith, B. 'Ontology'. In *Blackwell Guide to the Philosophy of Computing and Information*, edited by L. Floridi, 155–66. Oxford: Blackwell, 2003.
- Simon, H. A. *The Sciences of the Artificial*. Cambridge, MA: MIT Press, 1969.
- Slaughter, D. C., Giles, D. K., and Downey, D. 'Autonomous Robotic Weed Control Systems: A Review'. *Computers and Electronics in Agriculture* 61 (2008): 63–78.
- Shiffrin, R., and Schneider, W. 'Automatic and Controlled Processing Revisited'. *Psychological Review* 91 (1984): 269–76.
- Stahl, B. C. 'Emancipation in Cross-cultural IS Research: The Fine Line between Relativism and Dictatorship of the Intellectual'. *Ethics and Information Technology* 8 (2006): 97–108.
- Venkatesh, V. 'Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model'. *Information Systems Research* 11 (2000): 342–65.
- Veres, S. M., and Molnar, L. 'Documents for Intelligent Agents in English'. *Proceedings of the 10th IASTED International Conference* 674/122 (2010): 287.
- Visschers, V. H., and Siegrist, M. 'How a Nuclear Power Plant Accident Influences Acceptance of Nuclear Power: Results of a Longitudinal Study Before and after the Fukushima Disaster'. *Risk Analysis* 33 (2013): 333–47.
- Von Wright, G. H. *Humanismi Elämänasenteena*. [Humanistic stand to life] Helsinki: Otava [1981] 2007.

Passive Optical Sensor Systems For Navigation in an Unstructured, Non-Cooperating Environment

Christoph Sulzbachner, Christian Zinner, and Thomas Kadiofsky

Abstract

Due to the variety of applications of unmanned ground vehicles and unmanned aerial vehicles in civil and military domains, both autonomous and unmanned systems should be integrated into the specific domains. Studies showed important gaps in technological as well as legal aspects up to a respective integration of the systems into various application domains. In addition, current standards and regulations were not designed for either autonomous or unmanned systems. While in indoor scenarios assumptions can be made such as flat surfaces and perpendicular structures, in outdoor scenarios, which represent the vast majority of applications, such assumptions are not possible. For outside operations, structural maps are required that represent the environment as a whole and are kept very up to date by accumulating and registering acquired sensor data over time to build a consistent model that reconstructs the topography of the surface. Based on this surface, high-level algorithms can be applied, e.g., to find regions that are accessible by vehicles or usable airspace. This chapter gives an overview of passive optical sensor technologies and presents recent results in multispectral stereo vision for navigating in an unstructured, non-cooperating environment.

Introduction

Due to the variety of applications, there is rich potential for integrating unmanned ground vehicles and unmanned aerial vehicles into civil and military domains. Studies and official roadmaps, however, show that there are important technological and legal gaps that hinder implementation. In addition, current standards and regulations were not designed for either autonomous or unmanned systems. Meanwhile, a standardisation process was started in the

majority of application areas such as the automotive and aviation industry⁽¹⁾ on the international to the national level, in order to build a common baseline of requirements. There, robust navigation and safety functions are key indicators to allow integration. Today's vehicles already have operator-assisting technologies that mainly aim to reduce operational risks. These are some first steps in developing technologies for unmanned systems.

One of the major challenges of increasing the applicability of unmanned and autonomous systems is improving their ability to operate in outside environments, where the systems have to cope with a wide range of environmental conditions, and systems cannot automatically rely on man-made structuring elements such as roads, buildings, markings, signs, etc. Operating autonomous and unmanned platforms in such unstructured outdoor environments requires several major functional components, many of which can be supported by passive optical sensor systems (POSS). This chapter provides an overview of existing sensor modalities, with a focus on passive optical measurement principles. The next section describes basic functions relevant for unmanned platforms that can be directly supported by POSS technology. Subsequent sections describe important methods that rely on POSS and that have a high relevance to their application in autonomous aerial and land vehicles. Finally, a list of ongoing regulative activities is presented, which could support the integration of autonomous land and aerial vehicles in the civilian and military domains.

Key Functions Supported by POSS

This section provides a brief overview of functions that are relevant for the guidance and navigation of autonomous systems, and rates them according to the relevance of POSS. For the functions with significant relevance for POSS it is implied that sensors are used to get spatial information – thus, either the sensors themselves or the further processing of raw sensor data yields three-dimensional (3D) information with a more or less explicit characteristic.

- **By-wire steering and propulsion:** 'X-by-wire' refers to technologies that control units using electrical or electro-mechanical systems to perform

1. SAE International 2014; ERSG 2013; Federal Aviation Administration 2013; Kalra, Anderson, and Waches 2009; <https://www.eurocae.net>, accessed 2 April 2015; <http://jarus-rpas.org>, accessed 2 April 2015.

vehicle functionalities such as steering. Sensor equipment, however, does not play an important role; thus the POSS relevance is low.

- **Determination of vehicle position and time derivatives (i.e., motion, acceleration):** this can be split into two separate functions:
 - Navigation and localisation: common systems are based on global navigation satellite systems, radio navigation, or inertial guidance; these functionalities can be negatively impacted (or even disabled) by various influences such as atmospheric disturbances or jamming. More recent approaches involve methods of visual self-localisation with on-board cameras, thus the POSS relevance is high.
 - Speed and ego-motion⁽²⁾ estimation: aside from tachometers, wheel odometry, pitot tubes, and others, a vision-based method called visual odometry is used primarily on land vehicles; thus the POSS relevance is high.
- **Determination of the vehicle attitude:** several measurement units are used, such as inertia measurement units, magnetic field sensors, mechanical and laser gyros, etc. Under some circumstances, visual odometry can also contribute here; hence the POSS relevance is medium.
- **Object detection and collision avoidance:** several technologies such as radar or laser are used for state-of-the-art safety functions such as pre-crash warning systems. Camera systems and a respective fusion with existing sensors play an important role. The POSS relevance is high.
- **Mapping of the environment:** especially outdoors, learning and modelling the 3D structure of the environment is essential for navigation. Sensors such as laser have restrictions and POSS can contribute here, thus its relevance is high.
- **Mission guidance, operator view:** most kinds of missions, even with highly automated platforms, deliver feedback to an operator. The most-preferred form of feedback is a natural live view. In principle, images captured by on-board cameras for navigation, obstacle avoidance, and other tasks can also be used for this task; hence the POSS relevance is high.

2. Ego-motion is defined as the 3D motion of a camera within an environment.

This incomplete list indicates that POSS are among the key technologies that enable autonomous systems.

Passive Optical Sensor Technologies

This section provides an overview of optical sensors, especially the classification of 3D sensor technologies in general and the different characteristics of POSS. Generally, optical sensors convert electromagnetic energy into electronic signals that can be further processed. The electromagnetic spectrum is the totality of all electromagnetic waves of different energies. It can be divided into various bands with similar properties, as shown in Figure 11.1.⁽³⁾ There, the visible spectrum is defined as approximately 400–700nm.

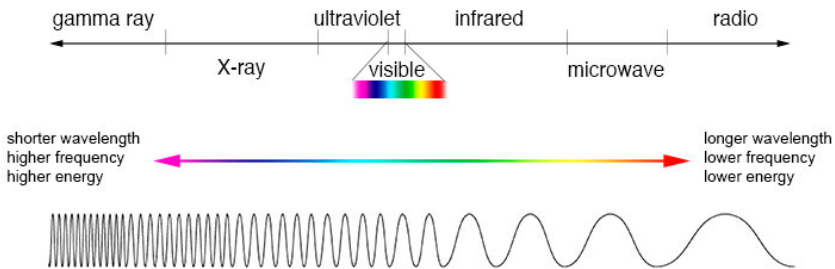


Figure 11.1. Electromagnetic Spectrum

However, depending on the optical sensor technology used, a specific range of the electromagnetic spectrum can be mapped onto a digital representation. Grayscale cameras cover the whole visible spectrum with one mapping; colour imagers disband the spectrum in colours; infrared imagers cover the infrared spectrum; and hyperspectral imagers divide the spectrum into a higher number of spectral bands, each with a specific spectral resolution.

3D Sensor Technologies and 3D Vision

3D sensors are designed to extract 3D information from ‘scenes’. Autonomously moving platforms in outdoor environments require a set of non-contact and non-destructive depth sensors, which are able to deliver

3. <http://imagine.gsfc.nasa.gov/science/toolbox/empectrum1.html>

spatial and temporal information at sufficient resolution. Figure 11.2 gives an overview of the state of the art of various sensing technologies, especially 3D

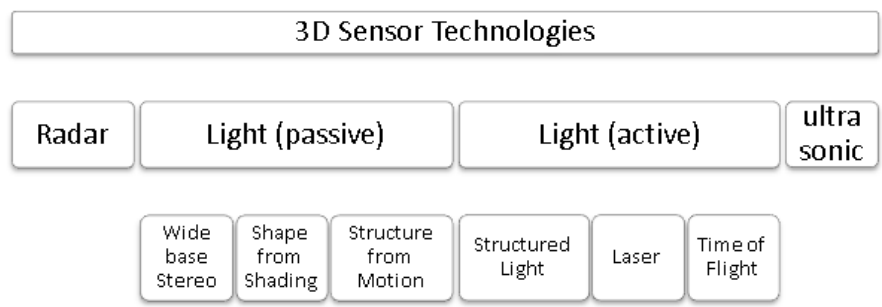


Figure 11.2. Overview of State-of-the-Art 3D Vision Technologies

sensor technologies.

In the automotive area, there are examples of on-board depth sensors for driver assistance systems such as adaptive cruise control systems, brake assistants, and piloted parking.

Optoelectronic 3D imaging can be classified into active and passive sensors. Active technologies emit signals into the environment to obtain distance information, while passive technologies rely only on the radiation received from the environment. Thus, active technologies have only a limited range in outdoor environments. Laser scanners are able to deliver relatively accurate 3D points; however, they require mechanical scanning facilities, which makes such devices complex and expensive, thus they are rarely found in consumer products. Time-of-flight sensors suffer from either high sensitivity to interfering ambient light and/or from high energy demands for the active illumination and their low lateral resolution. Although these are widely used in the scientific robotics community, they are clearly not applicable in outdoor environments with bright ambient light.

Passive 3D sensors allow the extraction of depth information from digital images captured from two or more different points of view. Conventional optical sensors produce only 2D projections of the 3D world. Methods that analyse image data from multiple cameras observing the same scene from different viewpoints allow the reconstruction of additional distance information.

Different viewpoints can be realised by either moving a single camera or by using more cameras that capture images simultaneously. Methods of 3D reconstruction using a single moving camera are still in the research stage, because the algorithms require a high computational effort and are thus usually not applicable under real-time conditions.

Canonical passive stereo vision methods are already used in several industrial applications, because real-time capable solutions are already state of the art, and they provide dense depth information at interactive frame rates. Stereo vision can deliver dense 3D data in real time, and it can cope with a wide range of ambient light conditions. 2D image information is also available. Common camera hardware is reasonably priced, there are no moving parts, and camera housings can be made very robust. The disadvantages of stereo vision are that depth resolution decays with increasing distance and untextured areas contain no depth information. Outdoor environments are usually highly textured, which is why passive stereo works better in outdoor than in indoor environments.

Optical Flow

Computation of the optical flow is the determination of the 2D motion vector field from two successive camera images. The vector field results from the movement of scene points and from camera ego-motion. Barron et al. distinguish between four groups: differential-based, region-based, frequency-based, and phase-based.⁽⁴⁾ Differential-based approaches use spatio-temporal derivatives of image intensity under the assumption that intensity is preserved during motion. Region-based methods use a similarity measure between the images to determine the displacement of pixels, similar to stereo vision. Frequency-based methods are based on velocity-tuned filters in the Fourier domain. Phase-based approaches assume that velocity is defined in terms of the phase behaviour of band-pass filter outputs. The fusion of the optical flow and the stereo vision results allows the computation of motion vectors in the 3D space. This approach enables new possibilities in terms of POSS.

4. Barron, Fleet, and Beauchemin 1994.

Visual Odometry

Methods of computing the position and orientation of a system by analysing the video input are called visual odometry. The basic assumption of these algorithms is that the major part of a scene is static; i.e., the geometric relations between points do not change over time. These relations, together with the properties and position of the camera, constrain how points in the world are mapped to images during the image formation process. Consecutive images are searched for characteristic points that can be recognised in these images. The motion of the camera is then computed from the displacements of these feature points within the images. This principle may be applied to a single camera or to a stereo camera setup. The drawback of monocular visual odometry is that the motion can only be determined up to an unknown scaling factor.

Visual odometry yields position and orientation updates in all six degrees of freedom of a rigid body motion. Hence, in order to get the trajectory of the camera, these position updates have to be integrated. Since each of the position updates is corrupted by small errors that accumulate during the integration process, the resulting trajectory suffers from drift – i.e., the accuracy of the

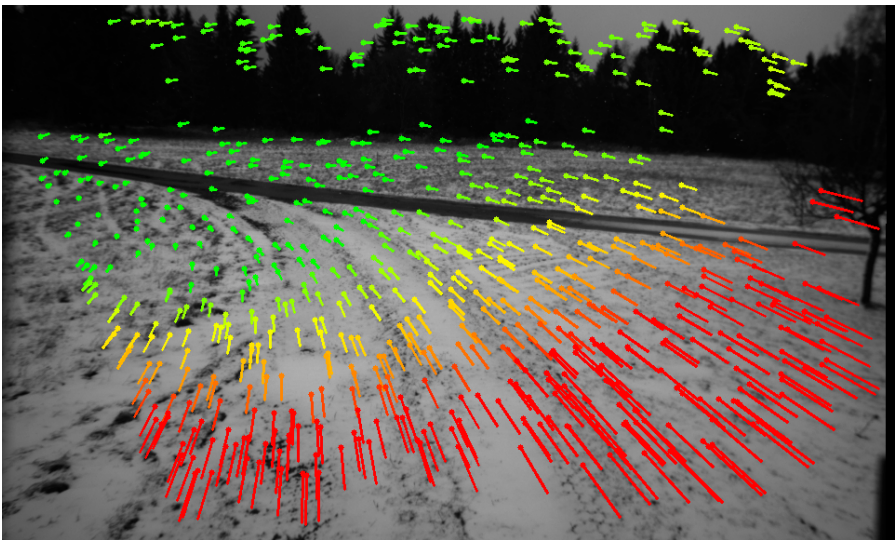


Figure 11.3. Visual Odometry: Motion Vector Field from Corresponding Feature Points in Consecutive Images

result decreases with the number of updates. A common countermeasure is to combine the results from the visual odometry with other sensors, e.g., an inertial measurement unit or a magnetic field sensor.

Outdoor Navigation in an Unstructured Environment

Navigation

The navigation process for moving a vehicle from location A to location B takes place on different levels. On a high level, only a very general course is planned, comparable to a satnav computing a route as lists of roads and junctions. In order to be able to follow this general course, the vehicle's path has to be planned in more detail. On this level, the trajectory is planned only for a short range. The results are commands such as 'move straight forward for x metres and then turn left by y degrees'. The basis of this path planning is the information about the vehicle's surroundings. Regions where the vehicle can move safely have to be identified and distinguished from those where the vehicle would harm itself or its environment. For this reason, vehicles are equipped with 3D sensors that can identify obstacles and determine how far away they are, and how fast they are moving. On the lowest level of the navigation process, concrete commands for the vehicle are derived to ensure it follows the planned trajectory. Finally, these commands (e.g., 20% acceleration, steer left, 50% brakes, or gear) are passed onto the actuators, and the vehicle moves along the planned path from location A to location B.

Human navigation is mainly based on visual perception. A short look at a scene is sufficient for making decisions about where it is safe to move. Computer vision systems try to imitate these capabilities, but are far from achieving the scene understanding of humans based on their experiences. Hence, the information about a vehicle's environment has to be converted into a form that can be processed and understood by a computer. Furthermore, single views generated by 3D sensors suffer from limited fields of view and resolutions, as well as occlusions. Therefore, obtaining a more complete reconstruction of a vehicle's surroundings requires combining multiple views from different positions and perspectives. The 3D data has to be processed and fused in a common coordinate frame, but it is typically captured relative to a sensor coordinate system. Hence, an important task is to recover the ego-motion

of the vehicle between consecutive measurements and to localise it within a common frame. Collecting and combining information over time and travelled distance imitates the human memory and keeps in mind regions, which cannot be observed by the sensors at a given time. As a result, the gathered information is most frequently represented by a map of the vehicle's vicinity.

Environmental Mapping

In indoor scenarios, assumptions can be made such as a flat planar floor and perpendicular structures; thus simple sensors might be sufficient to create 2D maps such as grid structures, with a binary classification indicating whether each cell of the grid is occupied by an obstacle or not. In outdoor scenarios, which represent the vast majority of applications, 2D maps are insufficient.

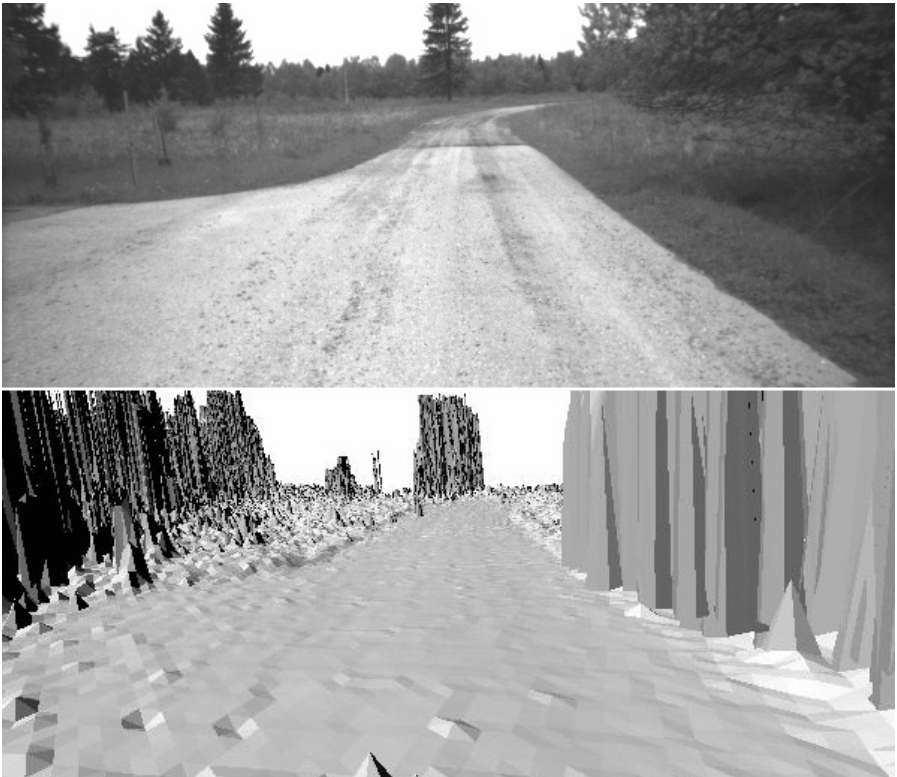


Figure 11.4. (Top): Camera Image; (Bottom): 2.5D Map Rendered from the Camera's Perspective

There, either 2.5D representations (such as elevation maps) are required, or 3D maps that represent the environment as a whole. Furthermore, in unstructured outdoor environments, a black-and-white classification in traversable and non-traversable regions is not optimal. There, all different shades of grey may occur. For instance grassland and a path are both quite flat and perfectly traversable. Nonetheless, there are differences like the speed that can be achieved on these surfaces, so the path should be preferred over grassland in a motion planning module. Hence, effort is also put into analysing not only the geometry of an environment, but also other properties that influence decisions about where to move a vehicle.

Dynamic Obstacle Detection

A collision-avoidance system for unmanned platforms requires both obstacle detection and collision avoidance that uses the information about the environment to extract accessible regions such as driveable regions or usable airspace. Several sensor approaches using active and passive sensor technology have already been discussed. In aerial applications, there are additional techniques using transponder systems, such as automatic dependent surveillance – broadcast. Depending on the airspace and vehicle, these kinds of systems are required in manned aircrafts. Using POSS, both static and dynamic approaches

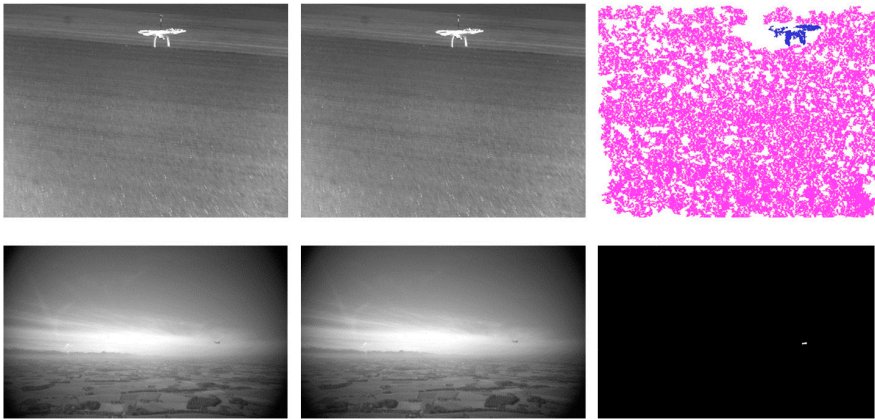


Figure 11.5. (Above): Stereo Vision Approach with both Left and Right Image with Dense Depth Map for Short-Range Environment; (Below): Dynamics Analysis Approach with Consecutive Images and Resulting Motion Analysis after Ego-Motion Compensation

can be applied to detect obstacles. Static approaches cover analysis techniques for finding obstacles in single or multiple frames, while dynamic approaches analyse the dynamic of the scene. Depending on the kind of vehicle, approaches have to be adapted, e.g., land vehicles can stop to further analyse the scene, while planes cannot.

Novel approaches for aerial obstacle detection are based on both approaches. For indoor and short-range environments using model planes with payload requirements, both monocular and stereo approaches can be applied, while in long-range environments using commercial unmanned aircrafts, payload requirements in terms of mass are marginal. Figure 11.5 shows the results of obstacle detection using a stereo vision sensor system and a monocular system that analyses the dynamics of the scene by compensating for the ego-motion.

Depending on the specific domain, the avoidance approach needs to fit the requirements. For land vehicles, the requirements are the road traffic act, etc., while aerial vehicles have to follow the aviation law, rules of the air, etc. that cover the domain-specific laws. For aerial applications, approaches can be rule based (e.g., implementing the rules of the air) or model based (e.g., turning right, stopping, or decreasing altitude as the vehicle (helicopter, plane) allows).

Regulations

To legally allow the use of autonomous and/or unmanned systems in real applications, the operations and legal requirements must be regulated. However, sensors that are able to capture the environment of autonomous systems are a vital precondition for realising such systems. Therefore, the abilities and limitations of the sensor modalities used, including POSS, play an important role in certifying (sub-)systems. Standardised methods for verifying optical and other sensor systems are still not fully established. This problem propagates up to the system level, which makes it hard to determine whether a given system fulfils a certain performance level that might be specified by legal regulations.

While various definitions and legal frameworks for the operation of autonomous systems have emerged already, further research and attention should be devoted to the related issue of proper verification methods for POSS.

Unmanned Ground Vehicles

Currently, there are no regulations for unmanned ground vehicles in place that allow driverless operation in the infrastructure. However, guidelines are in progress such as common definitions for autonomous driving or autonomous operation mode:

- National Highway Transportation Safety Agency levels (US): The NHTSA identified levels of automated vehicles covering no automation, function-specific automation, combined function automation, limited self-driving, and full self-driving automation.
- Bundesanstalt für Strassenwesen (Germany): The federal highway research institute distinguishes between assist, partially automated, highly automated, and fully automated; the degree of responsibility defines the level.
- SAE J3016: SAE International's On-Road Automated Vehicle Standards Committee distinguishes between the human driver and the driving system monitoring the driving environment on a higher level: no automation, driver assistance, partial automation, conditional automation, high automation, and full automation.

Common legal baselines include the Vienna and Geneva Conventions of 1968, which state that a driver must have full control of a vehicle at all times. This contradicts the integration of unmanned ground vehicles today. Various consortiums such as the United Nations Economic Commission for Europe – Transportation Division are working to update the Vienna Convention and harmonise vehicle regulations.

Unmanned Aerial Vehicles

In 2012, the European Commission established the European RPAS (Remotely Piloted Aircraft Systems) Steering Group (ERSG) to focus on the integration of unmanned systems in the European aviation system, including European Aviation Safety Agency, European Organization for the safety of air navigation, Single European Sky ATM Research Joint Undertaking, Joint Authorities for Rulemaking on Unmanned Systems (JARUS), and others. The ERSG developed a roadmap for the integration of civil remotely piloted aircraft systems into the European aviation system. This roadmap covers annexes on the regulatory approach, a strategic research plan, and a study of the societal impact of the challenges of RPAS integration. To achieve the full integration of

all types of RPAS into non-segregated airspace, the development of regulations in the domains of airworthiness, flight crew licensing, and air operations is essential. Due to EC Regulation 216/2008, RPAs with a maximum take-off mass above 150 kg are regulated by EASA, while RPAs below 150 kg are regulated at the national level by the national civil aviation authorities.

Currently, some European countries have national regulations in place, while others have prepared regulations. National regulations will likely be harmonised. The regulatory work plan has detailed regulatory improvements that should be performed by different stakeholders in different time frames. On the technological side, ERSO identified technological gaps in the safe integration of RPAs in the areas of: integration into ATM and airspace environments, verification and validation, data communication links (including spectrum issues), detect and avoid systems, and operational procedures, security issues, operational contingency procedures and systems, and surface operations (including take-off and landing). The strategic research and development plan described areas for future research programmes at the EU and national levels.

JARUS, a group of experts from national aviation authorities and regional aviation safety organisations, drafted an airworthiness code for light unmanned rotorcraft systems with a maximum certified take-off weight not exceeding 750 kg with the purpose of conventional helicopters. This certification specification covers both the airworthiness code and acceptable means of compliance. Draft deliverables of certification specification for light unmanned aerial systems or command and control links are also available.

References

- Barron, J, Fleet, D., and Beauchemin, S. ‘Performance of Optical Flow Techniques’
International Journal of Computer Vision 12/1 (1994): 43–77.
- European RPAS Steering Group (ERSG). *Roadmap for the Integration of Civil Remotely Piloted Aircraft Systems into the European Aviation System*, 2013.
- Federal Aviation Administration (FAA). *Integration of Civil Unmanned Aircraft Systems in the National Airspace System Roadmap*, 2013.
- Kalra, Nidhi, Anderson, James, and Wachs, Martin. *Liability and Regulation of Autonomous Vehicle Technologies*, Research Report UCB-ITS-PRR-2009-28. California PATH, 2009.
- SAE International. *J3016, On-Road Automated Vehicle Standards Committee*, 2014.

Toward Defining Canadian Manned-Unmanned Teaming (MUM-T) Concepts

Benoit Arbour, Matthew R. MacLeod, & Sean Bourdon

Abstract

As unmanned aerial vehicles (UAVs) mature from a capability developed during conflict into a standard element of nations' inventories, it is becoming more critical to consider how useful a UAV fleet might be across a wide spectrum of missions prior to their acquisition. As the reliability of automation increases, unmanned systems are given greater freedom of movement and action in light of the growing trust from the users. As a consequence, unmanned aircraft need no longer act in clearly different roles and airspace, and may instead closely support manned aircraft to achieve greater mission success in a wide range of traditional and non-traditional roles. The authors developed and executed a hierarchical decomposition method to evaluate the utility of a broad range of UAVs to a broad range of missions, specifically those that could be executed in concert with Canada's fighter and maritime patrol aircraft. Based on these results, initial advice is developed for both research and requirements staff on what types of UAVs will be most widely applicable to Canada's needs. Rather than providing an absolute ranking, the method is designed primarily to greatly narrow the decision space, so that the remaining options can be evaluated in more detail by technical and military experts. Finally, the chapter examines impact of the expected level of inter-aircraft interoperability (using e.g., NATO STANAG 4586) on the highly rated options, as well as the associated autonomy requirement.

Introduction

The use of UAVs in a military context is now fairly commonplace. Moreover, it is not unusual for manned and unmanned aircraft to collaborate while working toward a common military goal. In such a scenario, each aircraft may use its strengths to offset the other's weaknesses, with such pairings intended

to increase mission success. This may lead to specific unmanned capabilities being rapidly developed and fielded to fill specific gaps in manned aircraft capability. This is particularly common during wartime, when demands are at least partially driven by the needs of specific theatres matched with high technology readiness level (TRL) solutions, without an overall assessment of the long-term needs. Although notionally working together, the aircraft are generally disconnected from one another (often with the enforcement of strict altitude separation) or used as standalone capabilities without regard for the potential efficiencies introduced by the concept of manned-unmanned teaming (MUM-T).⁽¹⁾

When studying the concepts of integration and teaming, there is a tendency to automatically consider current commercial/military off the shelf (C/MOTS) solutions for both manned and unmanned aircraft. Although high TRL and a rapid deployment of capabilities are achieved, one can easily lock in to existing and temporary problems and solutions, or field sub-optimal systems that have been designed to solve different problems. On the other end of the spectrum, if C/MOTS solutions are not explored, research proposals may suggest new technology with a narrow focus to solve outstanding difficult niche issues – e.g., unmanned aerostats for remote Arctic surveillance – which may result in prototypes either falling short of (or not being generally applicable to) other requirements, or suggesting new technology to solve a wide class of problems with complex and expensive systems. Finally, given the costs associated with procuring new manned aircraft, focus is more often than not put on procuring new unmanned capabilities based on the unproven assumption⁽²⁾ that it will be faster and cheaper to acquire and operate UAVs. Pairing today's manned aircraft with available (or currently high TRL) UAVs may cause militaries to ignore mid- to long-term warfighting needs, while experimentation may fail to identify the right lessons for the aircraft of tomorrow by using restrictive assumptions or limiting experimental parameters.⁽³⁾ Lastly, experimenting

1. MUM-T is defined here as the use of manned and unmanned [aerial] vehicles in concert, i.e., when they can influence each other's course of action (Arbour, Bourdon, and MacLeod 2012).

2. The notion that UAVs are cheaper than manned aircraft is often used to justify their purchase, or at the very least to justify their use in dangerous situations as disposable assets. However, as noted in DOD (2013, 3): 'The size, sophistication, and cost of the unmanned systems portfolio have grown to rival traditional manned systems.'

3. 'Today's problems and their solutions may not solve the problems arising 20 years from now' (DOD 2013, 9).

with C/MOTS systems may also fail to recognise the synergies and efficiencies that could be achieved by developing MUM-T as a system with purpose-built sub-systems, rather than simply choosing combinations of pre-existing parts.

Complicating matters further is the fact that most militaries – particularly in nations such as Canada that have small to medium-sized armed forces – cannot afford to maintain many different types of equipment. It is thus important for decision makers in both research and acquisition to be presented with a comprehensive assessment of what types of platforms may be useful across a wide array of missions before rushing to acquire or develop specific solutions. Otherwise, developing a manned-unmanned team part by part for the problem(s) found in a specific theatre increases the risks of ending up with several niche assets, with little to bring to the next conflict.

In the current fiscal context many countries are facing, efficiencies must be found in order to ‘do more with less’. This problem may sometimes be resolved through the use of capability-based planning in order to evaluate the capability – or effect – one is trying to achieve in a military action in order to plan future fleets. However, the capabilities achieved by MUM-T may often be misunderstood. In particular, the combined effects generated by teaming aircraft may not be properly accounted for; rather, they may be seen as distinct contributors to mission goals whose value is limited by the capabilities they individually provide. Solutions engineered to seamlessly work in a cooperative fashion may therefore be overlooked in favour of capabilities one assumes can be integrated in the future, introducing potential financial and mission risk.

While teamwork is a natural means of overcoming problems in many settings of everyday life, the concept of MUM-T, with multiple aircraft acting as one (rather than multiple aircraft operating as a set of parts) has not been widely embraced. As a team, in the true sense of the term (i.e., where each individual works together to achieve a common goal), each can influence the other with its actions in order to modify a pre-defined and agreed-upon course of action. In a MUM-T concept, this can be achieved by much more than through the manned aircraft obtaining direct control of the UAV’s flight and airframe. NATO STANAG 4586⁽⁴⁾ defines five levels of interoperability (LOIs) starting at UAV control through a third party (LOI 1) up to full control

4. NATO 2012.

of the UAV within the manned cockpit (LOIs 4 and 5). Performing the same mission at a different LOI would result in the mission unfolding differently, and with varying degrees of success. Using the right LOI for the mission at hand is a challenge in itself.

It is clear that each LOI may necessitate a different UAV, or UAV type, which may not be a C/MOTS UAV built as a stand-alone system. But it is often forgotten that each LOI may also necessitate a different definition, or level, of autonomy on the UAV's part. By purchasing UAVs with built-in autonomy (from C/MOTS), defence departments implicitly agree with industry assumptions regarding the UAV's usage, rather than having industry design the right UAV for their needs. The autonomy level used for MUM-T experimentations may thus be improper for the expected LOI to be achieved, or workarounds must be put in place to offset the autonomy's limitations, thus further reducing the realism of the trials.

In the case study summarised herein,⁽⁵⁾ the authors were asked to examine the use of manned and unmanned aerial vehicles in concert, i.e., when they can influence each other's course of action, including whether there was any advantage to integrating them for use together, rather than treating them as separate fleets.⁽⁶⁾ Although a meaningful decomposition of the manned-unmanned problem space and a suitable assessment methodology were put in place, the MUM-T problem in the Canadian context is far from being resolved. General conclusions for the best UAV type, and the best UAV role, in support of two types of manned aircraft have been reached, but the interaction and autonomy levels necessary to achieve the perfect teaming, all the while ensuring fiscal viability and UAV multi-role capabilities, have yet to be pursued.

The intent of this chapter is to share Canada's experience in assessing the utility of UAVs in MUM-T operations, and to discuss how a similar approach may be used to explore the utility of acquiring or developing autonomous vehicles more generally. First, a description of how to decompose the problem space will be provided. This is followed by a brief overview of the mathematical

5. Although initially restricted to distribution within defence departments, the initial report on the case study has since been fully released under Access to Information request A-2014-00254 (National Defence and the Canadian Armed Forces 2014). See also Arbour, Bourdon, and MacLeod (2012).

6. Elements of this approach were also used in a workshop on the potential threat posed by the use of unmanned air/surface/sub-surface vehicles in the maritime environment (Haché 2013).

underpinnings of the work.⁽⁷⁾ Next, some conclusions from the original study will be discussed, leading to an assessment of possible follow-on work. Finally, conclusions relevant to the policy maker are drawn in the hope that this will encourage a more broad-ranging thought process when considering requirements for autonomous system research, development, and acquisition.

Structuring the Problem

The Canadian case study under the Manned-Unmanned Aerial Vehicle Interaction (MUAVI) project examined how useful the MUM-T concept would be for enhancing the ability of both the Canadian maritime patrol aircraft (MPA) and the Canadian fighter aircraft to achieve mission success. This included the current platforms (CP140 Aurora and CF188 Hornet, respectively) and their potential replacements. This chapter focuses on the underlying question of how UAVs can make a fighter aircraft more effective or efficient at its tasks when the MUM-T concept is employed, although will discuss the results of both analyses. In particular, this chapter explores possible MUM-T combinations and identifies the most useful avenues, while addressing lingering questions regarding the meaning of manned-unmanned interactions and UAV autonomy.

Many factors must be considered in this question. Both the MPA and the fighter are expected to accomplish a breadth of missions, especially in smaller militaries with limited fleets, in which an aircraft may need to perform duties that extend beyond its original purpose and capabilities. There is also a wide variety of UAVs – encompassing every class of manned aircraft as well as some that could never house a human occupant – each with their own strengths, weaknesses and predefined expectations vis-à-vis their employment. This makes for a large number of possible manned-unmanned pairings.

Structuring the problem space has been accomplished by developing a process to decompose the manned *platforms* – MPA and fighter – and UAV *types* (described below) into evaluable components. To start, the expected usage of each platform was defined in terms of its set of *missions*, as defined in the appropriate national documentation,⁽⁸⁾ including any expectations that

7. For a more generalised treatment of the method, see Bourdon, Arbour, and MacLeod (2014).

8. Canada's MPA and fighter missions were taken from each of their statements of operational requirement and operating intent.

evolved from there, or may evolve, for their eventual replacements.⁽⁹⁾ Next, each mission was decomposed into a set of *tasks* required to complete the mission. For example, the set of tasks for offensive missions tends to be fairly linear, following a kill chain analogy; however, tasks may also be concurrent. No mission or task prioritisation is assumed in the structure. The importance of each mission to the platform's expected usage, and the criticality of each task to the successful completion of the mission, were rated by appropriate subject matter experts (SMEs). Then, a set of *roles* that a UAV could perform in support of each platform was generated in order to assess the utility that an ideal UAV performing that role would bring to each of the platform/mission/task combinations. Finally, the ability of each UAV *type* to perform each of those roles was assessed at a high level.

The problem space decomposition described above yields a hierarchy of ratings; each level is logically combinable with the others, as shown in Figure 12.1.

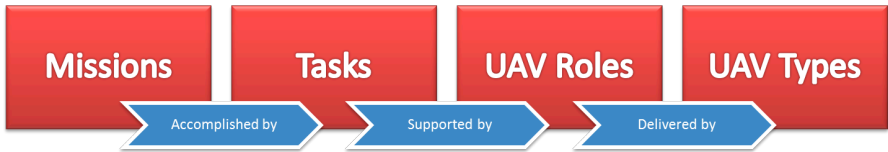


Figure 12.1. Linking Mission Requirements to Capability Delivery Options

The granularity at which each level of the hierarchy is defined can be varied depending on the analysis at hand; many different decompositions may arguably be suitable. In general, a good decomposition seeks to strike the appropriate balance between competing requirements at each level. Specifically, a decomposition is useful if it includes all important elements and is specific enough to distinguish between elements where appropriate, without introducing a

9. This may seem inappropriate in light of the discussion in the Introduction regarding the fact that the manned and unmanned aircraft form a team, rather than the sum of parts. However, it was deemed acceptable at such an early stage to evaluate how the MUM-T concepts would enhance the current and planned Royal Canadian Air Force (RCAF) capability in performing its current set of missions, and how the MUM-T concepts would change the way the RCAF does business. Adding new missions to the RCAF was out of the scope of the study.

level of detail that would make the workload unmanageable. This hierarchy was found to be granular enough to be usable and to show differences between the various options, but not so granular that it was unwieldy for the SMEs to provide all of the required ratings.

In order to ease the flow of the text, the exposition found below will focus on the fighter fleet, with appropriate MPA inclusions as necessary.

Missions and Tasks for the Fighter in the Canadian Context

In the Canadian context, the fighter⁽¹⁰⁾ missions and tasks have been broadly characterised as shown below. Most of the definitions closely follow the official NATO definitions:⁽¹¹⁾

- **Domestic Air-to-Air (A/A):**⁽¹²⁾ Ensure and maintain effective control over the Canadian territory, airspace, and maritime approaches including any role as defined within the NORAD agreements. The following tasks make up this mission:
 - **Transit:** travel from the base up to the edge of the area of interest (AOI);
 - **Detect:** search and detection;
 - **Intercept:** closing with an entity;
 - **Identify:** classification and/or identification of entities;
 - **Deter:** non-lethal action(s) influencing entities;
 - **Engage:** disabling of an entity; and
 - **Battle Damage Assessment (BDA):** assessment of the damage to an entity after an engagement.
- **Expeditionary A/A:**⁽¹³⁾ Comprises any situation in which the fighter force is required to deliver aerospace effects in a deployed role outside of Canada. The tasks associated with this mission are the same as for the domestic A/A mission.

10. No notion of how the UAV can support the fighter is assumed at this point. It is also important to remember that these are existing manned missions, and were not specifically chosen based on whether they could be enabled by using MUM-T concepts.

11. NATO 2014.

12. A/A is not an accepted NATO term, but the intent is reflected in NATO's air defence: all measures designed to nullify or reduce the effectiveness of hostile air action (NATO 2014).

13. Ibid.

- **Air-to-Ground (A/G):**⁽¹⁴⁾ Comprises any situation in which the fighter force is required to deliver ground effects.⁽¹⁵⁾ The tasks associated with this mission are the same as for the domestic A/A mission.
- **Air-to-Surface (A/S):**⁽¹⁶⁾ Comprises any situation in which the fighter force is required to deliver surface effects.⁽¹⁷⁾ The tasks associated with this mission are the same as for the domestic A/A mission.
- **Escort:** Involves the escort of one or more friendly unit through an area where they may be at risk. The friendly unit may be air, land or sea based. The following tasks make up this mission:
 - **Transit:** travel from the base up to the edge of the AOI;
 - **Detect:** search and detection;
 - **Determine Intent:** assessment of approaching entities;
 - **Deter:** non-lethal action(s) influencing entities;
 - **Engage:** disabling of an entity; and
 - **BDA:** Assessment of the damage to an entity after an engagement.
- **Search and Rescue (SAR):** The SAR mission involves the fighter as either the primary SAR asset or a support to a SAR asset in the search and rescue of distressed persons. Combat SAR is out of scope in the Canadian context. The following tasks make up this mission:
 - **Transit:** travel from the base up to the edge of the search area;
 - **Detect:** search and detection;
 - **Rendez-vous/track:** moving toward the unit(s) to be rescued and maintaining contact; and
 - **Rescue:** picking up stranded unit(s) or dropping a SAR payload.
- **Reconnaissance:** Includes the requirement to collect and transmit data on enemy or other targets of interest, including behind enemy lines. The following tasks make up this mission:
 - **Transit:** travel from the base up to the edge of the AOI;

14. A/G is not an accepted NATO term, but the intent is similar to NATO's antisurface air operation against ground, i.e., surface targets not at sea.

15. Note that domestic A/G is not sensible outside of training or an extreme wartime context, which A/G as envisioned here would cover.

16. A/S is not an accepted NATO term, but the intent is reflected in NATO's antisurface air operation: an air operation conducted in an air/sea environment against enemy surface forces (NATO 2014).

17. Similar to A/G, it was felt that engagements within Canada's waters would be limited to wartime scenarios and could be considered as part of this mission.

- **Data Capture:** recording of sensor data;
 - **Data Dissemination:** sending data to another friendly unit or to headquarters; and
 - **Detection Avoidance:** avoiding detection by enemy units.
- **Training:** This mission includes both the fighter crew being trained, and the fighter being used in the training of another platform. It refers to improving the existing training of aircrews, not to any additional training that would be required to interact with UAVs. The general concept is that a UAV could either take the place of a (potentially more expensive or unavailable) Blue unit that the fighter needs to be able to operate with, or simulate a Red unit to provide dissimilar air combat training. The following tasks make up this mission:
- **Simulate Blue:** simulate a friendly unit in a training scenario; and
 - **Simulate Red:** simulate a hostile unit in a training scenario.

Following the missions and tasks development, SME ratings were sought to determine the importance of each mission to the overall fighter usage and the importance of each underlying task to mission success. The ratings were then reviewed and discussed in order to obtain collective agreement. Those ratings can be found in Table 12.1 (3 = critical, 2 = important, and 1 = marginal).⁽¹⁸⁾

UAV Roles

The UAV roles are defined as the possible interactions the UAV may have with the manned aircraft on a MUM-T mission. For the current case study, eight roles have been defined and have been deemed sufficient to encompass the possible interactions within the fighter mission set. The eight roles are:

1. **Sensor:** acting either as an advance or complimentary sensor, feeding information to the manned aircraft that otherwise could not be obtained or which the manned aircraft does not have time to obtain;
2. **Refueller (Fuel):** exchanging fuel from an aircraft containing a large amount of jet fuel, the refueller, to another aircraft while in flight;

18. The ratings presented in Table 12.1 denote the consensus of the experts consulted during this process. For a given application of the method, a different scheme, such as averaging individual ratings, could also be employed. The authors chose consensus to stimulate discussions that would ensure that no one had mistakenly overlooked any key considerations in developing their individual ratings. For more information about missions, tasks, and associated ratings, see Arbour, Bourdon, and MacLeod (2012).

3. **Weapons Delivery (WD):** an armed UAV capable of launching ordnance at enemy units;
4. **Communications Relay (Comms):** a UAV serving as a node within a network for the purpose of exchanging data, information and commands over a wireless link;
5. **Decoy:** a Decoy UAV is designed to attract enemy weaponry or distract the opponent's attention, in order to allow other aircraft to either proceed undetected or, at the very least, unharassed by enemy units;
6. **Electronic Attack (EA):** For the purposes of this study, EA is separated from the wider category of electronic warfare (specifically, electronic support measures will be considered part of the sensor role). An EA UAV emits electro-magnetic energy to overwhelm, confuse, deceive or otherwise render ineffective the radar system of an enemy entity and/or its operator;
7. **Kinetic Employment (KE):** involves the use of the UAV itself as a weapon. Note that a UAV with the sole purpose of being used kinetically is deemed to be a weapon, not a UAV. Consequently, the KE role must not be the primary role of the UAV under study; and
8. **SAR Payload (SAR):** includes a number of measures to assist the 'Rescue' portion of SAR such as dropping supplies (e.g., food, medicine) or SAR technicians, possibly followed by extrication for transport to an appropriate location.

Each role can be complemented by a listing of characteristics that would need to be met by an aircraft (UAV or otherwise) performing the role. Such a list can be used to further bound the scope of the role, while also helping to determine how well each UAV type will perform the role. The SME-assessed utility of each role in each of the previously described fighter mission/task combinations is found in Table 12.1 (3 = very useful, 2 = useful, 1 = marginal, 0 = n/a).⁽¹⁹⁾

19. Because every UAV role is evaluated against every task, an 'n/a' rating can be used as a valid option in cases where the role cannot support the task: it was thought that attributing a 'marginal' rating whenever the role does not have a logical place within the task would unfairly penalise the role under any rating roll-up scheme. For more information about the ratings, including a listing of the UAV characteristics that have been deemed important in support of each role, see Arbour, Bourdon and MacLeod (2012). In the missions or rating importance scale, an 'n/a' was unnecessary as the mission or task could simply be dropped from the list.

Table 12.1. Canadian Fighter Missions, Tasks and Roles SME Ratings

Mission (Importance Rating)	Task (Importance Rating)	Sensor	Fuel	WD	Comms	Decoy	EA	KE	SAR
Domestic A/A (3)	Transit (3)	1	3	0	2	0	0	0	0
	Detect (3)	3	2	0	1	1	0	0	0
	Intercept (3)	2	0	0	1	0	0	0	0
	Identify (3)	3	0	0	2	0	0	0	0
	Deter (3)	2	0	3	2	3	3	2	0
	Engage (3)	3	0	3	3	2	2	2	0
	BDA (2)	3	1	0	3	2	0	0	0
Expeditionary A/A (3)	Transit (3)	2	3	0	2	2	0	0	0
	Detect (3)	3	2	0	1	1	0	0	0
	Intercept (3)	2	0	0	1	2	2	0	0
	Identify (3)	3	0	0	2	1	0	0	0
	Deter (3)	2	0	3	2	3	3	1	0
	Engage (3)	3	0	3	3	2	3	2	0
	BDA (2)	3	1	0	3	2	0	0	0
A/G (3)	Transit (3)	2	3	0	2	2	0	0	0
	Detect (3)	3	2	0	1	3	0	0	0
	Intercept (3)	2	0	0	1	2	2	0	0
	Identify (3)	3	0	0	2	3	0	0	0
	Deter (3)	2	0	3	2	3	3	2	0
	Engage (3)	3	0	3	3	2	3	2	0
	BDA (2)	3	1	0	3	2	0	0	0
A/S (2)	Transit (3)	2	3	0	2	1	0	0	0
	Detect (3)	3	2	0	1	1	0	0	0
	Intercept (3)	2	0	0	1	1	2	0	0
	Identify (3)	3	0	0	2	1	0	0	0
	Deter (3)	2	0	3	2	3	3	1	0
	Engage (3)	3	0	3	3	3	3	2	0
	BDA (2)	3	1	0	3	3	0	0	0

Mission (Importance Rating)	Task (Importance Rating)	Sensor	Fuel	WD	Comms	Decoy	EA	KE	SAR
Escort (2)	Transit (3)	3	3	0	2	2	2	0	0
	Detect (3)	3	2	0	1	1	0	0	0
	Determine Intent (3)	2	0	0	2	1	0	0	0
	Deter (3)	2	0	3	2	2	3	1	0
	Engage (3)	3	0	3	3	2	3	2	0
	BDA (2)	3	0	0	3	2	0	0	0
SAR (1)	Transit (3)	1	3	0	2	0	0	0	0
	Detect (3)	3	2	0	3	0	0	0	0
	Rendez-vous (3)	1	0	0	1	0	0	0	0
	Rescue (3)	3	0	0	2	0	0	0	3
Reconnaissance (2)	Transit (3)	2	3	0	2	0	0	0	0
	Data Capture (3)	3	2	0	2	1	0	0	0
	Data Dissemination (3)	0	0	0	3	0	0	0	0
	Detection Avoidance (2)	3	0	0	2	2	0	0	0
Training (3)	Simulate Blue (3)	3	2	0	1	3	2	0	0
	Simulate Red (3)	3	2	1	1	2	3	0	0

It is important to note that due to the nature of the questions asked to the SMEs, the ratings found in Table 12.1 are a statement on how useful the role is in an idealised sense, i.e., in a perfect world where the role performed by a UAV would be possible. Such ratings do not take into account several factors including, but not limited to, the following:

- The existence of a manned platform already performing the role (e.g., refuelling). The fact that a manned platform is capable of performing a role, or part of a role, does not impact the usefulness of a UAV in a MUM-T context also performing that role. The costs associated with having redundant capabilities do not impact UAV utility.
- The added workload to the crew of the manned aircraft or ground-based controllers. It is understood that a pilot (or crew, for a larger aircraft) will require cognitive power to interact with an unmanned partner. Similarly, ground-based controllers will need to be dedicated to the MUM-T concept. The difficulties encountered by either do not affect the usefulness of UAVs

in certain roles, as it was assumed from the onset that the appropriate level of autonomy could be achieved to alleviate those difficulties.

- Many other items such as the legalities, extra training, costs, manned aircraft modifications, etc. associated with having a UAV perform each role. These do not affect the utility of the UAV, but will certainly affect the feasibility and 'bang for the buck' of procuring such UAVs.
- The current capability of the manned-unmanned pairing in this context may lead to future research and development to add a capability, including the feasibility of incorporating UAV autonomy that is sophisticated enough to execute several complex functions.

In principle, there is no reason why someone applying this method to their own problem could not take these factors into account when developing the ratings, since the ratings are an evaluation of capabilities tied to the assumptions determined at the onset of the study. Consequently, most reasonable assumptions that are clearly articulated to the SMEs prior to establishing the ratings could be incorporated into the ratings.

As noted earlier, it is clear that any one of the factors enumerated above will at the very least impact any procurement process, and in many cases affect the successful completion of a MUM-T operation. However, due to the structure of the MUAVI project under which the current case study was developed, it was deemed necessary to use the first phase of work to evaluate the utility of possible MUM-T solutions in order to reduce the problem space by taking out the lesser-valued (in a perceived utility sense) manned-unmanned pairings. Each of the factors enumerated above was to be inspected in further phases (e.g., technical feasibility and the pilot workload assessment⁽²⁰⁾ was the purview of MUAVI Phase 2 while a cursory look at legal aspects⁽²¹⁾ was done in parallel with MUAVI Phase 1).

20. Pavlovic, Keefe, Fusina ND.

21. Barrett 2012.

UAV Types

For the purposes of the current analysis, a UAV has been defined as a ‘powered, aerial vehicle that does not carry a human operator and can fly autonomously or be piloted remotely’, with the caveat that ‘[a]mmunition, projectiles and missiles are not UAVs’ – in accordance with the Canadian Defence Terminology Bank.

UAVs are very numerous, and assessing each individual UAV in a MUM-T context would be completely impractical. Consequently, the authors sought to group UAVs into types with similar characteristics (to parallel the description of UAV roles in which desired UAV characteristics in performing the role were originally enumerated – not shown in the present case study). However, all of the extant UAV categorisations known to the authors appeared too restrictive; some UAVs were difficult to categorise or seen as ‘special cases’. Consequently, the authors designed simpler UAV groupings and avoided describing the precise delineation between each UAV type in order to avoid creating ‘special cases’. Hence, the vast number of known (and possible) UAVs was categorised into six generic types, which are not explicitly defined in order to be as inclusive as possible (some UAVs may fit into more than one category):

1. **Rotary Wing:** for example, the Northrop Grumman MQ-8 Fire Scout and the Boeing A160 Hummingbird;
2. **Fighter:** for example, unmanned versions of full-scale fighter aircraft such as the CF188 Hornet, F-35 Lightning II, and the F-22 Raptor;
3. **Airliner:** for example, unmanned versions of full-scale commercial aircraft such as the Boeing 737 and military versions of similar aircraft such as the CC130 Hercules and the CP140 Aurora;
4. **High/Medium Altitude Long Endurance (HALE/MALE):** for example, aircraft such as the General Atomics MQ-9 Reaper, the Northrop Grumman RQ-4 Global Hawk, and the CU170 Heron;
5. **Airship:** for example, all lighter-than-air and hybrid aircraft such as the Lockheed Martin High Altitude Airship (HAA) and Northrop Grumman’s Long Endurance Multi-Intelligence Vehicle (LEMV) concepts; and
6. **Small:** for example all tactical, mini, or micro UAVs such as the AeroVironment RQ-14 Dragon Eye, the Israel Aerospace Industries Mosquito, the CU161 Sperwer, and the CU167 Silver Fox.

In order to make an informed assessment of UAV capabilities vis-à-vis each role, a quick evaluation of each UAV type's typical capabilities was conducted using six categories: lift capacity (weight), lift capacity (size), speed, endurance, stealth, and manoeuvrability. Each UAV type is rated high, medium, or low capability for each category, as shown in Table 12.2.

Table 12.2. UAV Capabilities

	Rotary Wing	Fighter	Airliner	HALE/ MALE	Airship	Small
Lift Capacity (Weight)	Medium	High	High	Low	High	Low
Lift Capacity (Size)	Medium	Medium	High	Low	High	Low
Speed	Low	High	Medium	Medium	Low	Low
Endurance	Low	Low	Medium	High	High	Low
Stealth	Low	Medium	Low	Medium	Low	High
Manoeuvrability	Medium	High	Medium	Medium	Low	Medium

The UAV type characteristics as shown in Table 12.2 are typical characteristics that do not encompass any 'special case'. Using the UAV type characteristics as listed in Table 12.2 and the role definitions, each UAV type's capability to perform each role was assessed (see Table 12.3: 3 = very capable, 2 = capable and 1 = lacking).

Table 12.3. UAV Type against each Role Rating

	Rotary Wing	Fighter	Airliner	HALE/ MALE	Airship	Small
Sensor	2	3	3	3	3	3
Refueller	1	2	3	1	1	1
Weapons Delivery	2	3	3	3	2	1
Comms Relay	2	2	3	2	2	1
Decoy	2	3	3	3	1	3
EA	1	3	3	3	2	3
Kinetic Employment	2	3	3	2	1	2
SAR Payload	3	1	2	1	2	1

As shown in Table 12.3 through the sensor role, UAV types with very different characteristics may have been rated equally in terms of their ability to perform a specific role due to the nature of their employment: e.g., small UAVs are ‘very capable’ as close-up sensors, while airships are ‘very capable’ as persistent sensors. UAV types have thus not been penalised given the operational context in which they may perform the role; they were rated according to their expected usage in that specific role.

Data Analysis

Having gathered the necessary data, the perceived utility of UAVs in a MUM-T concept can be analysed. With each level of the problem space hierarchy (as shown in Figure 12.1) having been rated by appropriate SMEs, it is possible to ‘roll up’ the scores into something meaningful to the analyst and the client. For this purpose, the authors have developed the Hierarchical Prioritisation of Capabilities (HPC) method⁽²²⁾ allowing analysts to roll up the ratings and obtain results at any level of the hierarchy. The HPC starts with Table 12.4 as a generic representation of the rating breakdown from missions to roles. The table gives a sense of the combinatorial explosion that can be encountered if too complex a structure is adopted.

22. Bourdon, Arbour, and MacLeod 2014. A review of the HPC method closely following the work in Bourdon, Arbour, and MacLeod (2014) is presented in this section.

Table 12.4. Ratings for Roles Relative to Manned Aircraft Missions

Mission Name	Mission Rating	Task Name	Task Rating	Utility Rating		
				Role 1	...	Role k
A	w_1	A ₁	$w_{1,1}$	$w_{1,1,1}$		$w_{1,1,k}$
	
		A n_1	w_{1,n_1}	$w_{1,n_1,1}$		$w_{1,n_1,k}$
B	w_2	B ₁	$w_{2,1}$	$w_{2,1,1}$		$w_{2,1,k}$
	
		B n_2	w_{2,n_2}	$w_{2,n_2,1}$		$w_{2,n_2,k}$
...
M	w_M	M ₁	$w_{M,1}$	$w_{M,1,1}$		$w_{M,1,k}$
	
		M n_M	w_{M,n_M}	$w_{M,n_M,1}$		$w_{M,n_M,k}$

To compute aggregate ratings, the ratings were converted using the exponential function in order to have differences in ratings more closely represent the order of magnitude differences that were expressed in the descriptions of each level of the rating scales. In general, this better expresses the potentially large relative qualitative differences between ratings, even between consecutive numbers. In this fashion, a rating of 2, for instance, carries a weight of a^2 while a rating of 3 has a weight equal to a^3 , where a is the base for the exponential function. In order to facilitate interpretation of the results, the weighted sum is re-normalised into the same rating scheme as the inputs by using the logarithm with base a .

Given the ratings in Table 12.4 and the linkages shown in Figure 12.1, the HPC method uses hierarchical weighted sums to calculate measures of utility. The metric for the utility of role j in supporting mission i is expressed as:

$$u_{i,j} = \log_a \left(\frac{\sum_{p=1}^{n_i} a^{w_{i,p}} a^{w_{i,p,j}}}{\sum_{p=1}^{n_i} a^{w_{i,p}}} \right), \quad (1)$$

where $w_{i,p}$ is the rating of the p^{th} task relative to the i^{th} mission and $w_{i,p,j}$ is the rating of the j^{th} role relative to the p^{th} task relative to the i^{th} mission. The ratings of the importance of the task relative to the mission ($w_{i,p}$) are used as weights to better establish whether the roles have utility in accomplishing the tasks that are the greatest contributors to the overall missions. In this manner, the

formula assigns a greater score to roles that contribute at least a little to critical tasks than to roles that provide a great deal to tasks that are less important to the overall mission. This was deemed in line with operational thinking, in which the more important tasks must be performed to at least a minimum degree while lesser tasks may sometimes be omitted.

Similarly, the ratings can be aggregated one more level up the hierarchy. Specifically, when all the mission-specific ratings for a role have been computed, the overall utility of role j across all missions is calculated as:

$$u_j = \log_a \left(\frac{\sum_{p=1}^M a^{w_p} a^{u_{p,j}}}{\sum_{p=1}^M a^{w_p}} \right), \tag{2}$$

where the $u_{p,j}$ terms are calculated as in Equation (1) and w_p is the rating of the p^{th} mission relative to the overall set of missions. After determining each u_j , it is then possible for the analyst to determine which roles offer the greatest impact for each or all of the missions.

Given that the roles are delivered by UAV types (see Figure 12.1), it is possible to go one step further by assessing which UAV types have the greatest impact on the missions. Let's assume the generic SME ratings as shown in Table 12.5.

Table 12.5. Ratings of UAV Types in Fulfilling Specific Roles

	UAV Type			
	Type 1	Type 2	...	Type m
Role 1	$x_{1,1}$	$x_{1,2}$...	$x_{1,m}$
Role 2	$x_{2,1}$	$x_{2,2}$...	$x_{2,m}$
...
Role k	$x_{k,1}$	$x_{k,2}$...	$x_{k,m}$

The calculation of the rating of the utility of UAV type l to task j of mission i is as shown in Equation (3).

$$v_{i,j,l} = \log_a \left(\frac{\sum_{p=1, w_{i,j,p} \neq 0}^k a^{w_{i,j,p}} a^{x_{p,l}}}{\sum_{p=1, w_{i,j,p}}^k a^{w_{i,j,p}}} \right), \quad (3)$$

where $w_{i,p,j}$ is the rating of the p^{th} role relative to the j^{th} task in the context of fulfilling the i^{th} mission, and $x_{p,l}$ is the rating of the l^{th} UAV type's utility in fulfilling this role. The principal difference with previous equations is that terms that represent no utility for the role in fulfilling the task (i.e., where $w_{i,p,j} = 0$) are not included in the summation. This is to avoid penalising UAV types for not contributing a role that is not explicitly used in fulfilling the task.

From here, proceeding up the remainder of the hierarchy is much like before. The utility of a UAV type l in fulfilling mission i and across the full set of missions is given by:

$$v_{i,l} = \log_a \left(\frac{\sum_{p=1}^{n_i} a^{w_{i,p}} a^{v_{i,p,l}}}{\sum_{p=1}^{n_i} a^{w_{i,p}}} \right) \quad \text{and} \quad v_l = \log_a \left(\frac{\sum_{p=1}^M a^{w_p} a^{v_{p,l}}}{\sum_{p=1}^M a^{w_p}} \right), \quad (4)$$

respectively, where $w_{i,p}$, w_p and $v_{i,p,l}$ are defined as in Equations (1), (2), and (3). It is then possible for the analyst to determine the utility of a UAV type in fulfilling a mission or the overall set of missions.

Note that all formulae depend critically on the value of the base a . As discussed above, the choice of base allows the analyst to tailor the HPC method to the analysis at hand. While this number is difficult to precisely identify in practice, it was felt that a conservative estimate could be obtained in the context of the current case study. Specifically, in the present case study, the authors argued that it would be reasonable to ensure that seven ratings with a value of 1 do not surpass a single rating of 3 in importance. The number seven has not been randomly chosen: given the eight fighter missions outlined earlier, it was felt that given a single 'critical' mission – or 'very important' in the case of tasks – it would take at the very least all other missions to trump that single mission in perceived utility in a particular MUM-T pairing. Mathematically, this translates into the following inequality: $7a \leq a^3$, or $a \geq \sqrt[3]{7}$. Conversely, if enough missions or tasks have a rating of 2 ('important' or 'useful'), they should

collectively become more important than a single mission or task with a rating of 3, if they are in sufficient quantity. Again, a conservative estimate is to state that a single rating of 3 should not be greater in importance than seven ratings of 2 (following a similar logic as that for the 3's versus 1's). The corresponding inequality is $a^3 \leq 7a^2$, which in turn implies that $a \leq 7$. Combining the two inequalities restricts a to the range $[\sqrt{7}, 7]$. Since the two endpoint constraints are conservative, a value near the midpoint of the interval (i.e., $a = 5$) was chosen for the original analysis.

Case Study Results Using HPC Methodology

The results of the Canadian case study for the CF188 fighter aircraft using the HPC method and the Canadian SME ratings (see Tables 12.1 and 12.3) are shown in Table 12.6 for the missions versus roles and in Table 12.7 for the missions versus UAV types.

Table 12.6. Overall rating of UAV Roles against Canadian Fighter Missions

Mission	Sensor	Fuel	WD	Comms	Decoy	EA	KE	SAR
Domestic A/A	2.7	2.0	2.0	2.3	2.3	2.0	1.3	0.0
Expeditionary A/A	2.7	2.0	2.3	2.3	2.2	2.4	1.1	0.0
A/G	2.7	2.0	2.3	2.3	2.7	2.4	1.3	0.0
A/S	2.7	2.0	2.3	2.3	2.4	2.4	1.1	0.0
Escort	2.8	2.1	2.4	2.4	1.8	2.5	1.2	0.0
SAR	2.6	2.3	0.0	2.4	0.0	0.0	0.0	2.2
Reconnaissance	2.5	2.4	0.0	2.5	0.8	0.0	0.0	0.0
Training	3.0	2.0	0.7	1.0	2.7	2.7	0.0	0.0
Overall	2.8	2.0	2.1	2.2	2.4	2.4	1.1	0.1

Table 12.7. UAV Ratings against Canadian Fighter Missions

	Rotary Wing	Fighter	Airliner	HALE/ MALE	Airship	Small
Domestic A/A	1.9	2.8	3.0	2.8	2.6	2.7
Expeditionary A/A	1.9	2.9	3.0	2.8	2.5	2.8
A/G	1.9	2.9	3.0	2.9	2.4	2.8
A/S	1.8	2.9	3.0	2.8	2.5	2.8
Escort	1.9	2.8	3.0	2.8	2.5	2.7
SAR	2.2	2.5	2.9	2.5	2.5	2.4
Reconnaissance	1.9	2.5	3.0	2.5	2.4	2.3
Training	1.8	2.9	3.0	2.9	2.5	2.9
Overall	1.9	2.9	3.0	2.8	2.5	2.8

As can be seen, the sensor role has the highest expected overall utility by a wide margin. The Decoy and EA roles are second – keeping in mind that 0.4 is a large gap on the logarithmic scale with base 5 used. Niche roles that are well suited only for some missions can also be seen: for example, the decoy role for the A/G mission and the Refueller for the Reconnaissance mission.

Higher UAV type usefulness can be seen in the Airliner UAV – potentially a nod to the versatility of the current generation of these manned aircraft and their capability to take on many different payloads. The fighter type falls in second place, due to its commonality with the current Canadian fighter and its logical place as a wingman, followed the by HALE/MALE, due to the significant existing developments and the small type due to the potential for niche capabilities. Finally, the Rotary Wing UAV appears to be the least desirable as a complement to the Canadian fighter, potentially due to its slow speed and overt usage, compared to the small type with its slow speed and covert usage.

As an aside, the MPA analysis also showed that the sensor role scores much higher than the other roles, so much so that limited utility can be seen from the other roles in an overall sense. This result has been attributed to the assumed capability of the MPA to take on additional payloads and operators, thus minimising the need for a complementary platform. The problem as formulated by the client in this case was to look at how UAVs may complement existing platforms, but this result in itself is suggestive of the limitations of setting aircraft requirements independently rather than with a team concept in mind. The utility of roles to missions is highest (although not particularly high

in absolute terms) for the various Patrol missions (i.e., Anti-Surface Warfare Patrol, Anti-Submarine Warfare Patrol and Land Patrol). However, in general the utility is higher for the UAVs when coupled with the fighter than it is for the MPA, owing to the fighter's much more specialised role.

Case Study Observations

A list of trends, or specific results, of note from the original Canadian case study⁽²³⁾ is found below.

- The consistently high rating of sensor marks it as a very important role in support of both MPA and fighter. This is not particularly surprising, as it is really an enabler for every other mission.
- Other than sensor, none of the roles appeared especially relevant to the MPA. This is partially the case because the study looked only at cases in which the UAV would be used in concert with the MPA – whereas in many cases a UAV would be most applicable as a full-up replacement or alternative to the MPA.
- The decoy and EA roles appear to be a favourable option for pairing with the fighter. It should be noted that the expeditionary capabilities of the fighter force (to which these roles are applicable) are very important, but are not used that frequently. Both roles have their highest rating for training, allowing them to be useful during peacetime. The decoy role is also equally useful in an A/G capacity.
- The airliner type looks to be very broadly applicable, as it strikes a good balance of capabilities. This is perhaps not surprising, given the number of variants of, for example, the Lockheed CC130 Hercules currently in service.
- HALE/MALE aircraft appear quite capable, which is perhaps not entirely surprising given the research and development effort that has been expended in developing off-the-shelf models.
- From the data presented in Tables 12.1 and 12.6, niche UAV roles can be deduced such as SAR payload supporting a SAR mission, Refueller supporting a transit task, weapons delivery supporting an Engage

23. Arbour, Bourdon, and MacLeod 2012.

task, etc.⁽²⁴⁾ More specialised niche roles can also be envisaged such as Communications Relay in the Arctic region – although this conclusion needs the extra assumption (not presented in this chapter) that extra communications relay are necessary in the Canadian Arctic.

Interactions and Autonomy Considerations

Although the case study revealed interesting MUM-T combinations, the decision regarding which manned-unmanned pairing to pursue through acquisition is far from easy to make. The results of the Canadian case study (as shown in Tables 12.6 and 12.7 and in the Observations section above) highlight some key pairings with high perceived utility, but, as alluded to in the previous sections, high utility does not necessarily translate into the highest return on investment due to other constraints that must also be analysed. As the current case study looks solely at UAV capabilities in a MUM-T context, it is natural to leave any notions of costs and technical feasibility aside. However, UAV utility in a manned-unmanned formation is not solely determined by SME ratings and an HPC analysis.

Although the SME ratings found herein were produced under the assumption that an ideal UAV exists for the job being rated, no explicit discussion of how the MUM-T mission would be carried out occurred that renders the notion of ‘ideal’ fuzzy. As with any teaming efforts involving humans, determining the relationship and interactions between the parts (as well as the level of autonomy given to each part) will play a large role in the team’s success in achieving its goal. In a man-machine teaming, these notions are magnified as they must be defined from the onset in order to ensure that proper failsafe mechanisms are implemented, which implies that changes cannot be made on the fly to correct a teaming arrangement that has been deemed non-optimal for the situation: e.g., too little UAV autonomy might burden the manned crews, while too much autonomy may produce unpredictable behaviour that may limit the UAV’s usage in critical situations.

In order to be successful, a manned-unmanned team must act as one system in pursuit of its goal. Within this system, the interactions between the

24. Note that although obvious, any legitimate method must have these results rise to the top by the sheer nature of their links to specific tasks.

manned and the unmanned aircraft must be well defined in order to task each element appropriately and in order to pre-program the necessary elements in the UAV's software. The NATO STANAG 4586⁽²⁵⁾ offers five agreed-upon levels of interoperability (LOIs)⁽²⁶⁾ that characterise the type of interaction the manned aircraft may have with the UAV:

- Level 1: indirect receipt and/or transmission of sensor product and associated metadata, for example key-length-value (KLV) Metadata Elements from the UAV;
- Level 2: direct receipt of sensor product data and associated metadata from the UAV;
- Level 3: control and monitoring of the UAV payload unless specified as monitor only;
- Level 4: control and monitoring of the UAV, unless specified as monitor only, less launch and recovery; and
- Level 5: control and monitoring of UAV launch and recovery unless specified as monitor only.

As discussed in the Introduction, UAVs and manned aircraft currently fly within the same airspace, but only on rare occasions would they directly influence each other's action: a UAV gets imagery that, once dissected by imagery specialists, is redirected to the proper manned platform for action (LOI 1).

Although the Canadian case study has found that a Sensor UAV in support of a fighter performing an A/G patrol is very useful, using different LOI between the UAV and the fighter will cause the mission to unfold differently. Having the ability to monitor a UAV's payload in real time can allow the fighter to react more quickly than it otherwise would be capable of. By further having control of the UAV or its payload, the fighter can ensure the UAV will offer the imagery needed by the pilot without going through a third-party operator, which may introduce lag or errors in the system. However, increasing the LOI level, which may at first glance appear like the best thing to do, need not be a positive option:

25. NATO 2012.

26. Unless otherwise stated, LOI 3, 4, and 5 assume control and monitor.

- LOI 1: The fighter could be waiting to receive critical imagery from the UAV, or could see the UAV operators not respond to a request for specific imagery. This could easily result in mission failure.
- LOI 2: The fighter could see the UAV operators not respond to a request for specific imagery. Since the fighter need not wait to receive imagery; Level 2 may see an increase in mission success over level 1, but may still result in mission failure if specific requests are not met in a timely manner.
- LOI 3: The fighter will request and receive the necessary imagery from the UAV. This is the best-case scenario.
- LOI 4: The fighter will request and receive the necessary imagery from the UAV. However, the cognitive demands of flying the UAV may surpass the limits of the pilot's abilities. This may end up in mission failure.
- LOI 5: This is a more extreme case than level 4 and may also end up in mission failure.

The fighter's MUM-T A/G patrol LOI narrative – and the other missions' narratives that may end up with different conclusions for some LOIs – highlights the importance of defining the teaming arrangements ahead of developing a manned-unmanned team. Pairing the current manned platforms with high TRL UAVs designed as stand-alone systems evidently contradicts this conclusion.

There are many factors contributing to the continuing long tradition of avoiding a full evaluation of the MUM-T concepts. For one, the ease of simply adding a UAV to the battlefield and the success stories associated with LOI 1 – as opposed to not fielding UAVs, a sort of undefined LOI 0 – may convince decision makers that a solution has been found and the problem resolved. This may be inflated by the failure of fleet planning mechanisms, such as the capability-based planning previously discussed, to properly evaluate MUM-T capabilities that represent more than the sum of the parts. Or the costs and delays associated with the design, development, and acquisition of both manned and unmanned platform teams may appear unacceptable.

By failing to evaluate the gains provided by defining proper MUM-T arrangements, decision makers not only minimise the potential for extra gains in mission success but also disregard any impact on the fleets' mission set. The current problem's setup (looking at improvements in the current manned aircraft's mission set) or its extension (looking at improvements in the

current unmanned missions set) will lead to finding optimal aircraft pairings only. However, missions that are currently impossible with the manned or unmanned fleets may be possible using MUM-T concepts through the use of one aircraft to offset the other's limitations. Such a concept has already been discussed in the introduction as the reason for the influx of UAVs – along with the necessary LOI 1 – in the battlefield. However, pushing this notion further could open up many more opportunities that could influence how conflicts are fought and won: novel usage of assets may increase mission success, which in turn may shorten conflicts and decrease human casualties.

Conclusion

While utility, and interest, has been shown for the MUM-T concept in the Canadian context, understanding the proper teaming arrangements – such as the proper LOI and necessary UAV autonomy for the mission at hand – is still lacking. This situation has many roots. During ongoing conflicts there can be a real or perceived need to rapidly acquire and deploy forces, with an attendant default to C/MOTS options. While at least in the medium term it is inevitable that one or the other of the manned or unmanned aircraft will be a legacy fleet, which constrains the options, there is an opportunity to consider the broader applicability and potential of new acquisitions. There is also an opportunity for researchers and industry to look more broadly with their programmes, so that more generally useful unmanned aircraft are available as C/MOTS when the next acquisition cycle arises. In any event, fighting the next war with today's assets cannot reliably ensure future successes. MUM-T should be developed – or at least planned – as a system, rather than assembled parts, in order to develop the proper synergies and ensure the appropriate teaming arrangements. Failing this, an increase in mission success may occur, but at a much slower rate. Finally, the case study described herein followed a set of assumptions that could be changed in order to expand the problem space and assess the greater use of unmanned aircraft, such as evaluating the use of manned aircraft in support of unmanned missions or even assessing the MUM-T concept for missions that cannot currently be performed by either manned or unmanned aircraft. Moreover, such assessments of utility need not be limited to air systems: unmanned systems teaming with manned land or

maritime platforms can use similar techniques and arguments to maximise the long-term utility of the proposed MUM-T.⁽²⁷⁾

27. See, e.g., Haché 2013.

References

- Arbour, Benoît, Bourdon, Sean, and MacLeod, Matthew R. *Manned-Unmanned Aerial Vehicle Interactions Phase 1: Preliminary Concept Assessment*. DRDC CORA TM 2012-028. Ottawa: Defence Research and Development Canada, 2012.
- Barrett, Carrie. UAV Regulations Survey. Contract Report, DRDC CORA CR 2012-102, Defence Research and Development Canada, 2012.
- Bourdon, Sean, Arbour, Benoît, and MacLeod, Matthew R. 'A Method for Hierarchically Prioritizing Capabilities with an Application to Military Manned and Unmanned Aerial Vehicles'. *Journal of Applied Operational Research* 6/1 (2014):39–47.
- Haché, Chris. Maritime Remote Autonomous Systems Concept Development: MRAS Threats Workshop Quick-Look Report. Halifax, Nova Scotia: Canadian Forces Maritime Warfare Centre, 2013.
- National Defence and the Canadian Armed Forces. Completed Access to Information Requests 2014. Available at <http://www.forces.gc.ca/en/transparency-access-info-privacy/2014-completed-requests.page>, accessed 14 August 2014.
- North Atlantic Treaty Organization (NATO). NATO Glossary of Terms and Definitions (English and French). AAP-06, 2014.
- North Atlantic Treaty Organization (NATO). *STANAG 4586 Navy (Edition 3) - Standard Interfaces of UAV Control System (UCS) for NATO UAV Interoperability*. 9 November 2012.
- Pavlovic, Nada J., Keefe, Alan, and Fusina, Giovanni. 'Human Factors Issues in Airborne Control of Multiple Unmanned Aerial Vehicles'. DRDC Toronto TM 2013-098, Defence Research and Development Canada, Unpublished.
- US Department of Defense (DOD). *Unmanned Systems Integrated Roadmap: FY2013 - 2038*. 2013.

Merging Two Worlds: Agent-based Simulation Methods for Autonomous Systems

Andreas Tolk

Abstract

This chapter recommends the increased use of agent-based simulation methods to support the design, development, testing, and operational use of autonomous systems. This recommendation is motivated by deriving taxonomies for intelligent software agents and autonomous robotic systems from the public literature, which shows their similarity: intelligent software agents can be interpreted as the virtual counterparts of autonomous robotic systems. This leads to examples of how simulation can be used to significantly improve autonomous system research and development in selected use cases. The chapter closes with observations on the operational effects of possible emergent behaviour and the need to align the research agenda with other relevant organisations facing similar challenges.

Introduction

Modelling and simulation (M&S) is well known and often applied in NATO. Although mostly used in the training domain in the form of computer-assisted exercises, the NATO M&S Master Plan identifies five application areas that can capitalise on M&S: support to operations, capability development, mission rehearsal, training and education, and procurement.⁽¹⁾ This chapter therefore explores how M&S can best be used to support the various capacities of autonomous systems.

In their application-focused overview of M&S paradigms, Hester and Tolk describe the broad spectrum of M&S approaches:⁽²⁾

1. NATO 2012.

2. Hester and Tolk 2010.

- *Monte Carlo simulation* (the method of repetitive trials): This paradigm uses probabilistic models of usually static systems, and is used to evaluate systems that are analytically untraceable (such as those that cannot easily be described by mathematical functions).
- *Systems dynamics*: This paradigm is used to understand the behaviour of nonlinear, highly interconnected systems over time. It uses internal feedback loops, flows with time delays, and stocks and piles to model the system; systems are usually described using a top-down approach.
- *Discrete event simulation*: This paradigm is used for the dynamic simulation of systems in which the states are changing instantaneously when defined events occur. Event, time, and state change have to be defined precisely.
- *Continuous simulation*: This paradigm is used for the dynamic simulation of systems in which the states are changing continuously over time. They are normally described by differential equations that have to be approximated numerically.
- *Agent-based simulation*: The ‘agent’ metaphor uses agents as ‘intelligent objects’ that build a system from the bottom up, using agents to define the components of the systems. Agents perceive and act within their situated environment to reach their goals. They communicate with other agents following a set of rules. By adapting their rules to new constraints in the virtual environment, software agents can ‘learn’.

It is worth mentioning that system dynamics implements a typical top-down design approach, while agent-based models are more useful for building systems from the bottom up based on component descriptions. Discrete event simulation supports both approaches, but traditionally is used more often to implement top-down solutions.

While all M&S paradigms can provide some support to autonomous systems, the agent-based simulation paradigm is of particular interest, as autonomous systems reflect characteristics similar to those of software agents. Autonomous robotic systems⁽³⁾ are defined by the ability (e.g., by using integrated sensing, perceiving, analysing, communicating, planning, decision making, and acting/

3. The insights derived in this chapter are primarily based on the field of robotics, which is why the term ‘autonomous robotic systems’ is used. It is reasonable to assume that the results are generalisable, but a formal evaluation has not yet been conducted.

executing) to achieve assigned goals.⁽⁴⁾ Intelligent software agents are defined by their ability to perceive their situated environment; socially interact with other agents for planning and executing; act in their environment based on their programmed beliefs, desires, and intentions; observe the results; and adjust their actions based on these results.⁽⁵⁾

The research presented in this chapter not only shows that the characteristics of autonomous robotic systems and intelligent agents are similar, but also that the taxonomies are alike as well. This is important because it means that agent-based simulation methods can be used to support autonomous robotic systems in various application domains. Furthermore, observations of emergent behaviour typical of agent-based systems are likely to be observed in autonomous system populations as well.

Characteristics and Taxonomy

This section describes the characteristics and taxonomy of intelligent software agents, as they are used within agent-based modelling, and of autonomous robotic systems, as they are dealt with in this book. Its goal is to provide the researchers of both domains with a basic understanding of why it is pivotal for them to work together to maximise the benefit for NATO.

Intelligent Software Agents

This section is mainly derived from the contribution of Tolk and Uhrmacher to the seminal work of Yilmaz and Ören.⁽⁶⁾ The agent metaphor is a well-researched topic, but the results are distributed among a huge variety of research domains. The metaphor is based in various computer science areas – such as distributed systems, software engineering, and artificial intelligence – and has been strongly influenced by research results from disciplines such as sociology, biology, cognitive sciences, systems sciences, and many others. Although there are many definitions of agents, the following working definition provided by

4. Huang, Messina, and Albus 2003.

5. Yilmaz and Ören 2009.

6. Tolk and Uhrmacher 2009; Yilmaz and Ören 2009.

Tolk and Uhrmacher describes the main characteristics of intelligent software agents:⁽⁷⁾

- *The agent is situated, it perceives its environment, and it acts in its environment.* The environment typically includes other agents, other partly dynamic objects, and passive ones, which are, for example, the subject of manipulation by the agent. The communication with other agents is of particular interest in systems comprising multiple agents, as agents can collaborate and compete for tasks. This latter characteristic has also been referred to as ‘social ability’.
- *The agent is autonomous,* in the sense that it can operate without the direct intervention of humans or others; autonomy requires control of its own state and behaviour. Agents must be guided by some kind of value system, which leads to the often-used statement: ‘objects do it for free, agents do it for money’.⁽⁸⁾ In other words, an object always executes functions that are invoked, while agents can decide if (and how) they react to a request.
- *The agent is flexible,* which means it can mediate between reactive behaviour (being able to react to changes in its environment) and deliberativeness to pursue its goals. A suitable mediation is one of the critical aspects for an agent to achieve its tasks in a dynamic environment. An agent can act upon its knowledge, rules, beliefs, operators, goals, and experiences, etc. and to adapt to new constraints and requirements – or even new environments – as required. For example, new situations might require new goals, and new experiences might lead to new behaviour rules. In other words, an agent can learn. Furthermore, being mobile adds to its flexibility.

The following figure exemplifies these characteristics. It shows an intelligent agent in the centre of its environment. The agent perceives its environment, which includes other agents he can interact *with* and objects he can act *on*. He maps this perception to an internal representation, which may be incomplete,

7. Tolk and Uhrmacher (2009, 77). As a rule, intelligent software agents are virtual entities that exist in software programs. However, this technology is already applied commercially to support intelligent internet-based software agents that support the automatic updating of travel arrangements, recommend new products to customers, etc.

8. Jennings, Sycara, and Wooldridge 1998.

e.g., he only knows about one of the square posts and nothing about the triangular-shaped object. Objects can be static or expose dynamic behaviour, like the ball that will roll once it is kicked. Advanced agents may use simulation to act on their perception to support their decision-making process; for example, they may simulate each alternative and apply measures of merit that reflect their goals, desires, and beliefs onto the projected result and select the alternative with the highest expected value. An agent communicates with other agents and acts on the objects, such as kicking the ball in the direction that the other agent is running. If the plan does not work as expected, the agent will learn from his observations that this action does not lead to the desired outcome, and he will choose other options in the future. If something works well, he will use this strategy more often. Holland describes several learning algorithms that can be applied, and many more have been developed and successfully applied since then.⁽⁹⁾ Some agents use game theory approaches to select a mixed strategy based on the expected pay-offs of possible alternatives.⁽¹⁰⁾

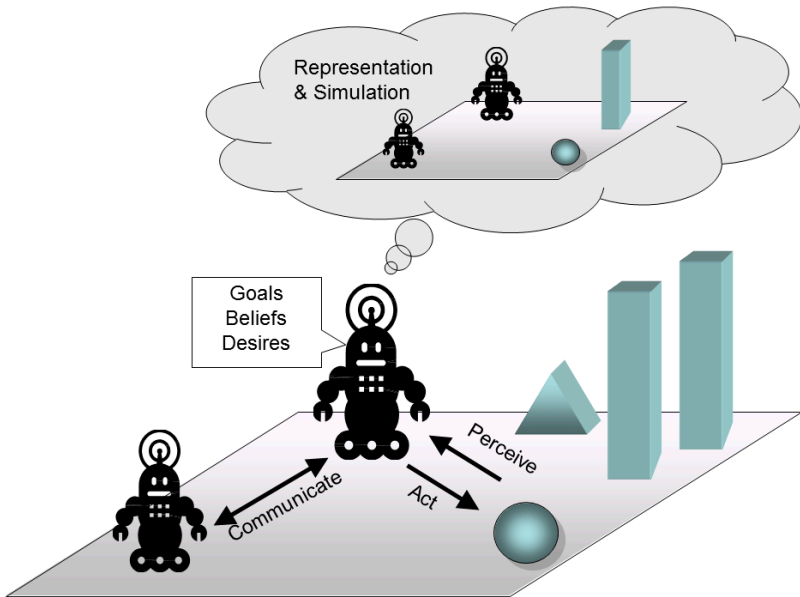


Figure 13.1. Intelligent Software Agents in the Situated Environment

9. Holland 1986.

10. Parsons and Wooldridge 2002.

In order to implement these characteristics, several architectural frameworks were recommended in the agent-based literature, many of which focus on the particular area of interest. Moya and Tolk 2007 composed the various ideas into a common architectural framework that captures the taxonomy of a single intelligent software agent.⁽¹¹⁾ This taxonomical structure was kept simple and adaptable in order to support as many viewpoints as possible. It identifies three external domains and four internal domains.

The three external domains comprise the functions needed within an agent to interact with the situated environment, which includes objects and other agents:

1. The *perception domain* observes the environment. Using its sensors, the agent receives signals from his environment and sends this information to the internal sense-making domain.
2. The *action domain* comprises the effectors. If the agent acts in his environment, the necessary functions are placed here. It receives the task to perform tasks from the internal decision-making domain.
3. The *communication domain* exchanges information with other agents or humans. If it receives information, it is sent to the internal sense-making domain. It receives tasks to send information from the internal decision-making domain.

The four internal domains categorize the functions needed for the agent to decide how to act and adapt as an autonomous object (see Figure 13.2):

1. The *sense-making domain* receives input (via sensors and communication) and maps this information to the internal representation that is part of the memory domain. These domains comprise potential data correlation and data fusion methods; data mediation capabilities; methods to cope with uncertain, incomplete, and contradictory data, etc.
2. The *decision-making domain* supports reactive as well as deliberative methods, as they have been discussed in this chapter. It uses the information stored in the memory domain and triggers communications and actions.
3. The *adaptation domain* may be connected with perception and action as well, but that is not a necessary requirement. The comprised function

11. Moya and Tolk 2007.

group updates the information in the memory domain to reflect current goals, tasks, and desires.

4. The *memory domain* stores all information needed for the agent to perform his tasks. It is possible to distinguish between long-term and short-term memory, and different methods to represent knowledge can be used alternatively or in hybrid modes.

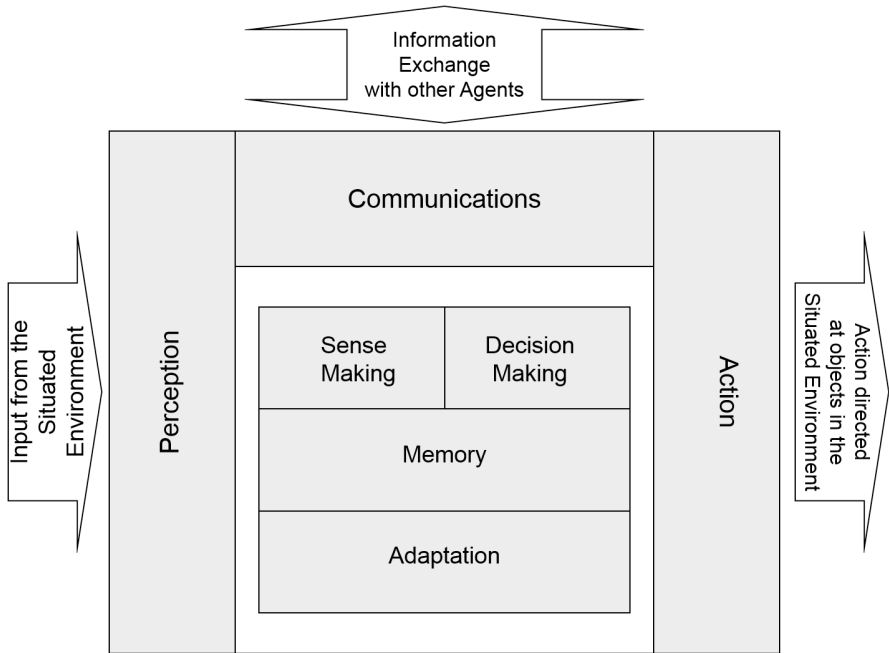


Figure 13.2. General Taxonomy of an Intelligent Software Agent

A complete agent taxonomy for agent-based simulation needs to reflect not only on the individual agents, but also on the characteristics of the situated environment as well as of the agent society. Wooldridge identifies five categories to characterize the environment for an intelligent software agent:⁽¹²⁾

1. *Accessibility*. The environment can be *accessible* or *non-accessible*. This category addresses how much of its attributes the environment exposes.

12. Wooldridge 2000.

This is different from the question of how the agent perceives what is exposed, and whether he can make sense of what he perceives.

2. *Determinacy*. The environment can be *deterministic or stochastic*. In deterministic environments, an action always has the same deterministic effect. In stochastic environments, this is not the case. For all practical purposes, the ‘real world’ can be assumed to be stochastic in nature.
3. *Reactivity*. The environment can be *episodic or sequential*. In episodic environments, the action is only relevant for the current episode. In sequential environments, an action may have effects in future states as well. This includes the idea of ‘effects of effects’ that often take some time to be exposed. The ‘real world’ is sequential in nature.
4. *Degree of change*. The environment can be *static or dynamic*. A static environment does not change during the evaluation period; a dynamic environment does. The real world is dynamic in principle, but in many practical applications is static for this particular application.
5. *Type of change*. The environment can be *discrete or continuous*. Furthermore, discrete environments can differ in the level of resolution, accuracy, and granularity. This category directly connects back to the modelling paradigm used to provide the environment for the software agents.

In their analysis of different collections of agents and how they act in society, Moya and Tolk identified the size (number of agents) and diversity (type of agents) as the driving categories in the literature dealing with agent societies.⁽¹³⁾ To cope with all observations, two additional categories – social interactions and openness – were proposed to characterize the agent societies built by intelligent software agents more generally.

- *Size*. The number of agents within the population can vary between large-scale numbers of several thousand agents down to a few agents. In some cases, only a single agent is used, although this is the exception.
- *Diversity*. The society can comprise agents that are all of the same type, building a homogeneous society, or agents of different types, building a heterogeneous society. It is worth mentioning that even agents of the

13. Moya and Tolk 2007.

same type can exhibit different behaviour depending on their state and initialization.

- *Social interactions.* The agents can either cooperate with each other or be in competition. In addition, all mixed forms are possible, such as coalitions that cooperate with each other but compete with others. It is also possible that agents are agnostic and that every agent follows its own objectives without any interaction.
- *Openness.* The agent society can be open or closed. In open societies, anyone can contribute agents and add them to the society. In closed societies, the number of contributors is limited by constraints. The contributors can be human, other agents, or systems. As before, various mixed forms are possible.

The resulting agent taxonomy, which reflects all characteristics of agenthood regarding the agent, the situated environment, and the agent society, is exemplified using the top-level categories in Figure 13.3.

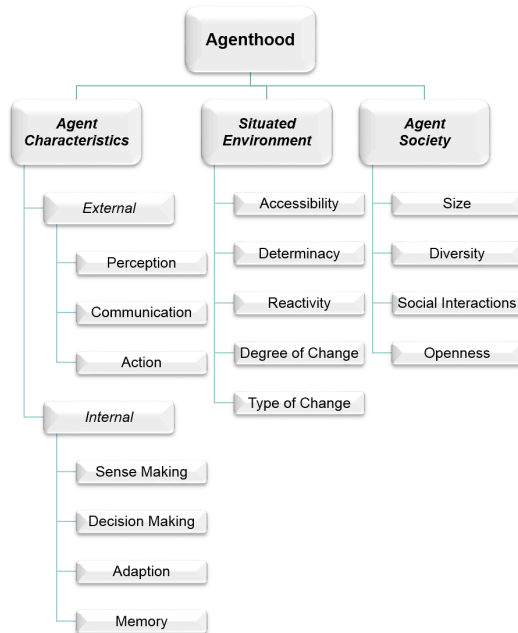


Figure 13.3. Taxonomical Components Describing Agenthood

The viewpoint of this summary of characteristics and taxonomy of intelligent software agents is biased toward the applicability of agent-based modelling methods in support of autonomous robotic systems, in particular to show the similarities.⁽¹⁴⁾

Autonomous Robotic Systems

The characteristics and taxonomical structures presented in this section are simplified to improve easier understanding of the mapping potentials between components of the agent-based simulation domain and the domain of autonomous robotic systems.

As with intelligent agents, a multitude of heterogeneous application domains and supporting disciplines contributed to the definition of autonomous robotic systems. A common understanding of autonomy is that the system has the capability to make decisions about its actions without the involvement of an operator, which also entails entrusting the system to make these decisions. This is far more than automation, which involves using control systems and information technology to reduce the need for human intervention within well-defined constraints.

The Autonomy Level for Unmanned Systems project by the National Institute of Standards and Technology defines autonomy as:

The condition or quality of being self-governing. When applied to unmanned autonomous systems (UAS), autonomy can be defined as UAS's own ability of integrated sensing, perceiving, analysing, communicating, planning, decision-making, and acting/executing, to achieve its goals as assigned by its human operator(s) through designed human-robot interface (HRI) or by another system that the UAS communicate with.⁽¹⁵⁾

Autonomous robotic systems are an example of UAS. Their characteristics are close to the working definition presented for intelligent software agents above. The main difference is that software agents are virtual actors in a virtual

14. For more detailed information, see Yilmaz and Ören 2009.

15. Kendoul 2013.

world, while robots are physical actors in the physical world. The following list has been compiled to clarify the similarities of intelligent software agents and autonomous robotic systems.

- *The autonomous robotic system is situated, it perceives its environment, and it acts in its environment.* The environment typically includes other autonomous robotic systems and objects that are subject to manipulation by the system. In scenarios with multiple robots, they can collaborate and compete for tasks. Therefore, autonomous robotic systems often have the ability to communicate with each other.
- *The robotic system is autonomous* in the sense that it can operate without the direct intervention of humans or others; autonomy requires control of its own state and behaviour. As a rule, it needs to have a plan that is often assigned by one or several human operators.
- *The autonomous robotic system is flexible.* If the observation shows that current actions do not lead to the desired effects, the autonomous robotic system can identify and execute alternatives. Advanced systems may even create a new plan together. Mixed strategies between immediate reactive behaviour and more time-consuming deliberate behaviour ensure that the system is safe in critical environments.
- *The autonomous robotic system is mobile*, in the sense that it can move in the environment within the physical constraints imposed on it. Eventually, additional rules of engagements may set more constraints than just the physical ones.

The resulting taxonomy presented here is a simplified aggregate of the ideas presented in Mataric's and Siegwart, Nourbakhsh, and Scaramuzza's seminal works on robotics and autonomous mobile robots.⁽¹⁶⁾ The following components comprise the taxonomy of an autonomous mobile robotic system:

- *The locomotion component* moves the system in its environment, and is constrained by the different degrees of freedom. These can be tracks or wheels, but also the rotors of a helicopter or quadrocopter, etc.

16. Mataric 2007; Siegwart, Nourbakhsh, and Scaramuzza 2011.

- *Actuator components* are moving parts of the autonomous robotic system, such as robot arms, sensors, antennas, etc., which are used in order to act on things, perceive better, etc.
- *Manipulation components* interact with objects and the environment. They grab, push, turn, and do whatever is needed to act on the environment to conduct actions according to the plan.
- *Sensor components* observe the environment. They are the eyes and ears of the robot. They can be passive or active. They can be as easy as switches operated by bumpers, or they can be lasers and sonars or complex cameras.
- *Signal processing components* are used to convert sensor signals into computable information. They are sometimes seen as components of the sensor, but Mataric points out that these components are also used to convert computed information into actuator signals. As such, they are sitting between the eyes, ears, and arms of the robot, and its brain.
- The *control component* is the 'brain' of the autonomous system. It makes the decisions based on the perception created from the input of the sensors and the plan the robot is following. As a rule, the control component is a computer.
- *Communication components* exchange information with other robots, as well as with humans, via HRI. The variety of communication components is as big as that of sensors, but they all serve the same purpose: allowing the control component of the robot to exchange information with other entities.
- Of critical importance are the *power supply components*, as they are the energy source for all actions. Usually, these are batteries or solar panels, but alternatives are possible as well, depending on the size and the tasks of the robot.

Figure 13.4 displays the components using the structure of the intelligent agent taxonomy as a guide.

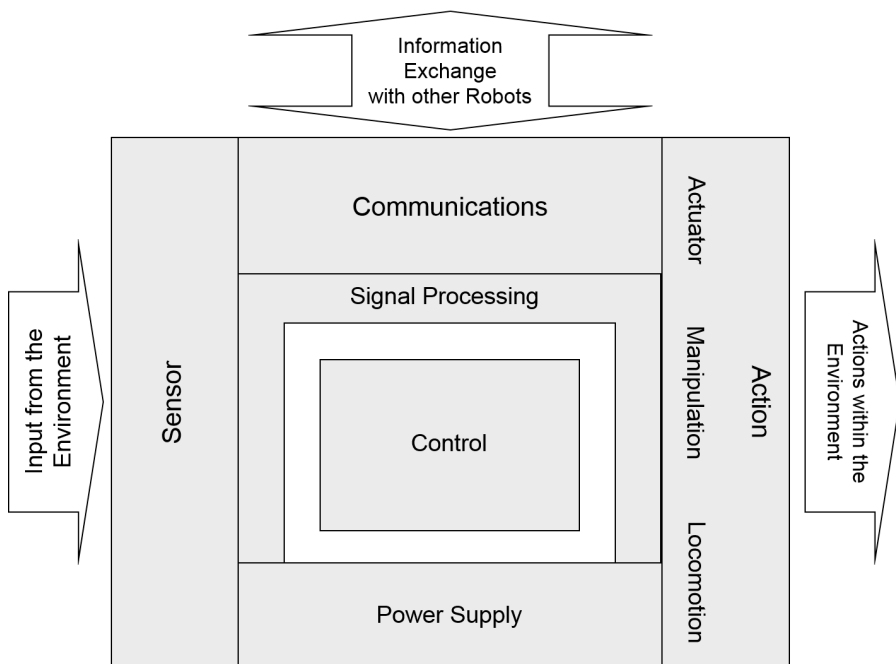


Figure 13.4. General Taxonomy of an Autonomous Robotic System

There are, without a doubt, many differences between agents and robots. Agents live in the virtual world, while robots live in the physical world. As such, robots have many tasks regarding locomotion, perception, localization, and navigation that the agents do not have to address in the virtual world. A lot of research has therefore been directed at better sensors, better locomotion and manipulator components, and other components that are necessary to make a robot work in the physical environment. As a result, the external domains are dealt with in greater detail in the robotics domain than in the agent domain. Many robotics practitioners even regard simulation as inferior, stressing that there is a huge reality gap between the needs of robotics and the contribution capability of simulation.

In contrast to such perceptions, the focus in this chapter shifts toward domains in which intelligent agent research already provides results that are directly applicable to improving autonomous robotic systems. The social ability of agents; the ability to learn; and the algorithms developed, implemented, and tested regarding sense making, decision making, and other components of the

interior domain can support more intelligent robotics behaviour. Furthermore, the simulated environment of agents is a safe and inexpensive test bed for the intelligent behaviour of autonomous robotic systems as well.

The taxonomical similarity between intelligent software agents and autonomous robotic systems allows the identification of these areas of mutual support. In the next section, several application domains are identified that can be immediately utilized to improve the behaviour of robots by applying methods from the agent paradigm.

Applying the Agent Paradigm in Support of Autonomous Robotic Systems

This section is neither complete nor exclusive. The objective is to give three examples of very different application domains showing the synergy of intelligent software agents, M&S, and autonomous robotic system research. The first example shows the synergy in the domain of test and evaluation, starting earlier in the procurement phase, where first testing is possible using only virtual prototypes of an envisioned autonomous robotic system. The second example shows how sense making and machine understanding, as used in intelligent software agents, can be used successfully to improve the sensing and perceiving activities needed in autonomous robotic systems as well. The third topic shows the relevance of research results in the domain of intelligent software agents for operational experts who are interested in using autonomous robotic systems.

Developing and Testing Autonomous Robotic Systems

Many simulation publications give examples of using simulation as a test bed to stimulate systems under test, some of them more than a decade ago, such as McKee.⁽¹⁷⁾ The US Army launched the Simulation and Modelling for Acquisition, Requirements and Training initiative to support this domain,⁽¹⁸⁾ which spawned several follow-on activities in the other services as well as in

17. McKee 1998.

18. Page and Lunceford 2001.

NATO. Neugebauer et al. present related work on a distributed test bed based on simulation services.⁽¹⁹⁾

Today, the idea of using simulation to provide stimulation for a system under test is well established and often applied. However, autonomous robotic systems are posing new challenges for the test-and-evaluation community. In addition to exposing a great variety of requirements from many stakeholders, the main challenge is that autonomous robotic systems operate in a dynamic and unpredictable environment, and onsite testing of the total feature set of a new system under realistic operational conditions is impractical.

Agent-based test environments can help address these challenges in two ways. First, they allow the behaviour of an autonomous robotic system to be tested using its virtual counterparts before it is implemented within the physical robot. An intelligent software agent can follow the same rule sets in the virtual environment that the autonomous robotic system would follow in the real environment. Its sensors can be simulated, as can its interactions with objects. This idea is not new, as documented by Akin et al. using the example of the RoboCup Rescue Robot and Simulation Leagues.⁽²⁰⁾ Many of these simulated rescue robots can be programmed in the same programming language and with the same tools that the real robots will be using later on. The Defense Advanced Research Project Agency Robotic Challenge uses the same approach. Aleotti et al. envisioned such an approach early on.⁽²¹⁾

The author supported the US Navy with research on the ‘Riverscout’, an unmanned surface water vehicle that demonstrates some autonomous behaviour as well. Figure 13.5 shows various simulation screen shots as well as the real system in a test. Some additional information has been published by Barboza.⁽²²⁾

The second way in which agent-based models support the testing is by providing a smart and adaptive test environment. The test environment is not just a script-driven stimulation provider that allows the researcher to systematically provide all sorts of inputs; it actually reacts meaningfully to the actions of the system. In other words, agent-based models realistically replicate the dynamic and unpredictable operational environment for the test. Furthermore,

19. Neugebauer, Nitsch, and Henne 2009.

20. Akin et al. 2013.

21. Aleotti, Caselli, and Reggiani 2004.

22. Barboza 2014.

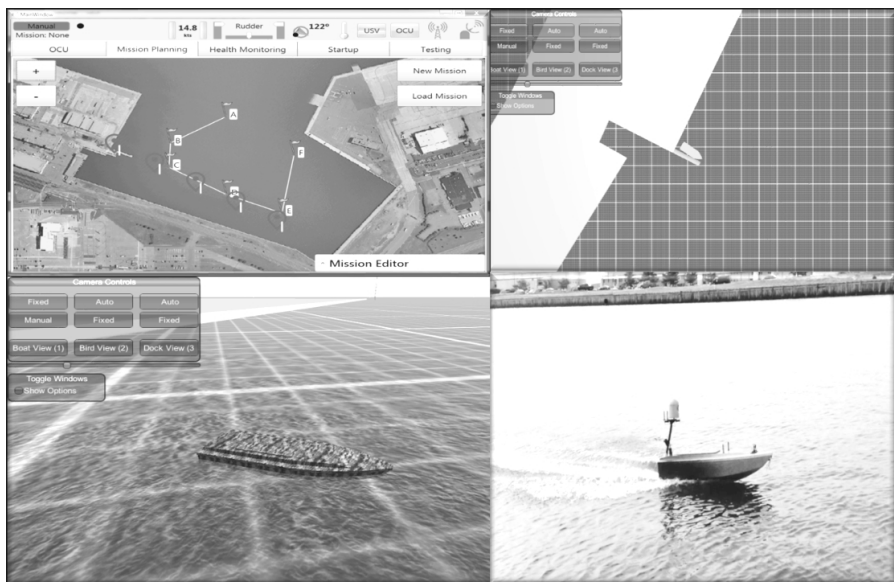


Figure 13.5. Simulation Support for the Unmanned Surface Vehicle 'Riverscout'

they offer the other autonomous robotic systems that are needed to provide the full operational functionality within a scenario realistically. Every system is represented by intelligent software agents. If the ideas of executable architectures based on operational specifications (as provided by common system architectures) are applied, as described in more detail by Garcia and Tolk, each intelligent software agent can take over the role of a system with the envisioned portfolio to provide the most realistic operational conditions for testing, even if not all systems of the portfolio are physically available.⁽²³⁾

A major challenge for agents representing human actors has been to elicit expert knowledge in a form that agents can use. Hoffman et al. showed that Applied Cognitive Task Analysis and critical decision methods allow for the creation of a cognitive representation for agents within the virtual environment.⁽²⁴⁾ Garrett conducted a series of experiments to show their applicability for

23. Garcia and Tolk 2013.

24. Hoffman, Crandall, and Shadbolt 1998.

intelligent software agents.⁽²⁵⁾ This research is highly relevant when autonomous robotic systems are supposed to be surrogates for human experts.

These last two paragraphs also emphasize that the behaviour of an intelligent software agent shall never be rooted in interpretations of a software developer, but driven by experts' insights and operationally validated artefacts. This insight becomes even more important when the methods are applied to implement intelligent behaviour for robots, as they act (and interact) in the physical world, and failures and wrong decisions may have dangerous – and even deadly – consequences. The autonomous robot system community should therefore carefully analyse the lessons learned from the intelligent software agent community.

The first systems using related technologies are already in operational-like use. The Control Architecture for Robotic Agent Command and Sensing project conducted by the Office of Naval Research evaluates swarm technology, which is a sub-set of agent-based technology, to control a set of autonomous surface vessels to protect selected ships. The sensor and software kit can be transferred between small vessels that are under human control, but follow simple rules that all contribute to a new capability to better protect ships. Similar approaches have been successfully tested for search and rescue operations.

Using Agent Methods to Implement Intelligent Behaviour

Even in their seminal book on autonomous mobile robots, Siegwart et al. explicitly state that the focus lies on mobility.⁽²⁶⁾ Comparing the two taxonomical structures presented in Figures 13.2 and 13.4 also shows that the focus of agents is the sense making and decision making to act meaningfully in the virtual world, while the focus of robots is more geared toward interacting with the physical world. Again, this represents a possibility to create synergy by bringing both worlds together and using intelligent software agent methods to enable autonomous robotic systems to expose the same degree of sense making, decision making, memory, and adaptation as agents do.

25. Garrett 2009.

26. Siegwart, Nourbakhsh, and Scaramuzza 2011.

An applicable lesson learned goes back to Zeigler’s work on how machines gain understanding.⁽²⁷⁾ This model will be explained here in a slightly modified form. Nearly 60 more types of machine understanding, many of them also applicable to this chapter, have been evaluated by Ören et al., and many have been successfully applied within intelligent software agents.⁽²⁸⁾ They provide a rich body of knowledge that the autonomous robotic system community can draw from. In order for a machine to understand, four premises have to be fulfilled, as they are visualized in Figure 13.6:

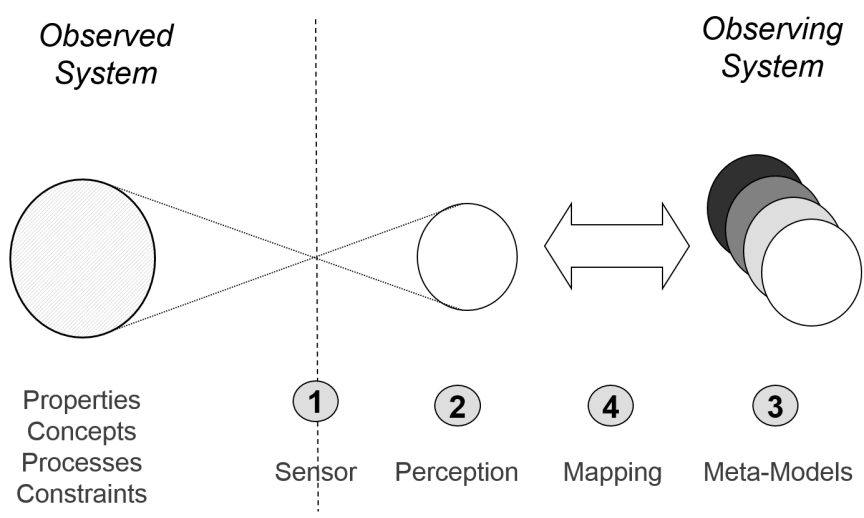


Figure 13.6. Using Meta-Models and Mappings in Support of Machine-based Sense-making

1. The machine must have sensors to observe its environment. The type of sensor defines which attributes can be observed and which properties and processes can be recognized if the target-noise ratio between the observed system and the situated environment is high enough.⁽²⁹⁾

27. Zeigler 1986.

28. Ören et al. 2007.

29. Some general sensor modelling constraints have been documented in Tolk (2012).

2. The machine has a perception of the system to be understood. The properties used for the perception should not significantly differ in scope and resolution from those exposed by the system under observation. They are often closely coupled with the sensor's abilities as well as with the machine's computational abilities.
3. The machine has meta-models of observable systems, which are descriptions of properties, processes, and constraints of the expected behaviour of observed systems. Understanding is not possible without such models. They can be understood as types of systems that may be observed in the situated environment.
4. The machine can map the observations, resulting in the perception of a suitable meta-model explaining the observed properties, processes, and constraints.

Understanding involves pairing the perception with the correct meta-model. If such a model does not exist, a sufficiently similar model can be used that explains at least part of the behaviour. The learning algorithms mentioned above can then be applied to adapt existing models or to create new models that can be applied to future observations.

This principle is not too far from how humans gain knowledge. When primitive peoples first make contact with higher civilizations, they often use familiar terms to address new concepts that are 'close enough'. Examples may be 'giant metal birds' when addressing airplanes, or 'giant locusts with human faces prepared for battle' when describing attack helicopters with pilots in their cockpits. The concept of a bird addresses the ability to fly, but one of the properties is very different from birds, as they normally are not made out of metal. Once enough information is collected, new concepts can be created to deal with the observed system, so that it can be recognized.

Another aspect of interest is the possibility of capturing desires and beliefs in machine-understandable form. Harmon et al. documented canonical structure to allow mimicking the behaviour of both single human beings and collectives.⁽³⁰⁾ Again, this knowledge can be applied to make autonomous robotic systems act 'more human' if this is in the objectives of the development.

30. Harmon et al. 2001.

Emergence and Operational Implications

Intelligent software agents are often connected with emergence. Maier distinguishes in the context of system-of-systems engineering between weak, strong, and ‘spooky’ emergence.⁽³¹⁾ Looking at this mainly from the systems perspective (not as an artificiality of the representing model or software), he defines the terms as follows:

Weak Emergence: An emergent property that is readily and consistently reproduced in simulations of the system, but not in reduced complexity non-simulation models. It can be understood through reduced complexity models of the system after observation, but not consistently predicted in advance.

Strong Emergence: An emergent property that is consistent with the known properties of the system’s components but which is not reproduced in any simplified model of the system. Direct simulations of the system may reproduce the emergent property but do so only inconsistently and with little pattern to where they do so and where they fail. Reduced complexity models or even simulations do not reliably predict where the property will occur.

Spooky Emergence: An emergent property that is inconsistent with the known properties of the system’s components. The property is not reproduced in any model of the system, even one with complexity equal to that of the system itself, even one that appears to be precisely simulating the system itself in all details.⁽³²⁾

A well-known example of the general emergence of system behaviour that results from simple rules between the implementing components is Schelling’s segregation model.⁽³³⁾ It has been implemented multiple times to show how people use very simple rules to select their neighbourhood. This decision is based on a preferred mix between two population groups. On the system level, segregation patterns emerge without any implicit or explicit formulation of such a property on the component level.

31. Maier 2014.

32. Ibid., 22

33. Schelling 1969.

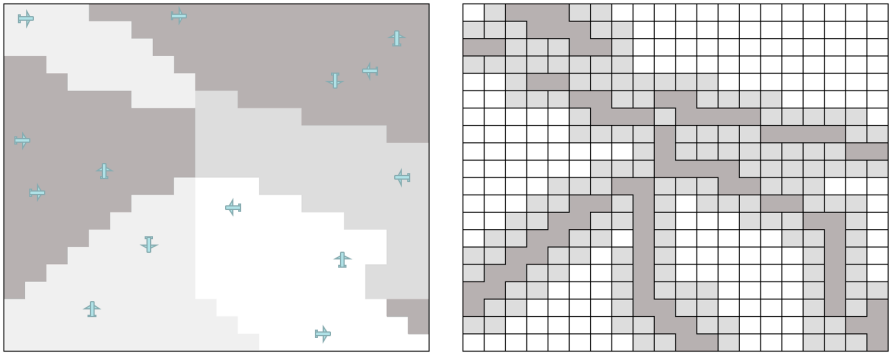


Figure 13.7. Emergent Observation Patterns and Gaps

Bonabeau's seminal paper on using agent-based modelling to represent self-deciding entities made a strong connection between emergent behaviour in the system and such entities.⁽³⁴⁾ While accepted in social systems, such emergent behaviour is often not wanted in technical systems, such as societies of autonomous robotic systems. Nonetheless, as the system composition is analogous to that of an agent-based population, emergent properties are more likely to be observed.

The following figure shows an example implemented in Netlogo, which is an open system developed by the Center for Connected Learning at Northwestern University.⁽³⁵⁾ The left figure shows six zones in which agents can move. The rule for the agents is to move around the zone using an equal distribution to cover all ground, but to avoid colliding with other agents in the same zone and to keep a safe distance from agents in a neighbouring zone. An operational interpretation could be drones in the air zones of coalition partners that want to observe their areas of responsibility. When counting the number of times an agent visits another agent, the right picture emerges: there are some islands close to the border between the areas of responsibility that are rarely visited, although no rule excluded these areas from observation.

In an operational context, such areas can easily become safe havens for activities that were supposed to be observed in the first place. While this example is trivial and easy to fix, it demonstrates the underlying challenge: how can a

34. Bonabeau 2002.

35. Tisue and Wilenski 2004.

commander support positive emergence that helps reach his objective while avoiding negative emergence that is counterproductive?

A first step is to raise awareness that emergence in systems comprising interacting autonomous robotic systems will occur. It is not an artificiality. It will happen, and it needs to be controlled in the context of the operational objectives. A second step is to conduct more research on whether – and to what degree – it is possible to engineer positive emergence into such systems and avoid negative emergence. Although initial results have been published, and some answer may even be found in the fundamentals of cybernetics as described by Ashby, the topic itself is still in its infancy.⁽³⁶⁾ Tolk and Rainey are making it a priority in their recommended research agenda.⁽³⁷⁾ The engagement of the autonomous robotic system community is highly encouraged, as the results of this research are directly applicable to societies of autonomous systems as well.

Summary

In the only recently released report on this topic,⁽³⁸⁾ internationally recognized experts recognised the perpetually increasing importance of modelling and simulation for the design, development, testing, and operation of increasingly autonomous systems and vehicles for a wide variety of applications on the ground, in space, at sea, and in the air. Section 4 of the report identifies the four most urgent and most difficult research projects:

1. behaviour of adaptive/non-deterministic systems;
2. operation without continuous human oversight;
3. modelling and simulation; and
4. verification, validation, and certification.

All four research domains are at least touched on this chapter: (1) agents are virtual prototypes that can be used to evaluate the behaviour, (2) establishing robust rules and procedures allows the operation to be conducted without human oversight, (3) the use of modelling and simulation is an important area of research, and (4) verification, validation, and certification are an

36. Ashby 1963.

37. Tolk and Rainey 2014.

38. NRC 2014.

important aspect that justifies the use of agent-based test beds. This shows not only the necessity of related research, as recommended in this chapter, but it also demonstrates the potential of fruitful collaboration in this domain with other research agencies. Many civil agencies and government organisations are interested in these questions. The results of their efforts should be observed, as it is likely that many of their key findings can help answer the urgent NATO research questions.

The domain of intelligent software agents has an enormous potential to enrich the research on autonomous robotic systems. It can (and should) be applied in the domain of procurement:

1. test and evaluation – testing robotic behaviour in a virtual environment before the robot is built;
2. providing a flexible and intelligent test bed for autonomous systems, as traditional test and evaluation methods are insufficient to test systems' autonomous characteristics;
3. using established methods and algorithms to enable the learning and adaptability of robots, as intelligent software agents are well known for their ability to learn and adapt to new situations (and the topological similarity shown in this chapter implies that many ideas can be mapped and reused); and
4. operational support – from building awareness of new challenges and demonstrating new capabilities to evaluating emergent behaviour in the operational context.

Both communities should actively engage to develop synergisms and potentially multi-disciplinary collaboration with NATO, which has the necessary structures for common research and presentation in place. Hopefully, this chapter will help establish a common research agenda and encourage mutual benefit from results that are already available.

References

- Aleotti, J., Caselli, S., and Reggiani, M. 'Leveraging on a Virtual Environment for Robot Programming by Demonstration.' *Robotics and Autonomous Systems* 47/2 (2004): 153–61.
- Akin H.L., Ito, N., Jacoff, A., Kleiner, A., Pellenz, J., and Visser, A. 'RoboCup Rescue Robot and Simulation Leagues.' *AI Magazine* 34/1 (2013): 78–86.
- Ashby, W.R. *Introduction to Cybernetics*. Hoboken, NJ: Wiley, 1963.
- Barboza, J. *Conversion Challenges of an Autonomous Maritime Platform: Using Military Technology to Improve Civilian Security*. Paper presented at the MODSIM World Conference & Expo 2014, Hampton, VA, 15–17 April 2014.
- Bonabeau, E. 2002. 'Agent-based Modeling: Methods and Techniques for Simulating Human Systems'. *Proceedings of the National Academy of Sciences of the United States of America* 99/3 (2002): 7280–7.
- Garcia, J.J., and Tolk, A. 'Executable Architectures in Executable Context enabling Fit-for-Purpose and Portfolio Assessment'. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* (2013), 10.1177/1548512913491340.
- Garret, R.B. 2009. 'A Method for Introducing Artificial Perception (AP) to Improve Human Behaviour Representation (HBR) Using Agents in Synthetic Environments'. Ph.D. Dissertation. Norfolk, VA: Old Dominion University.
- Harmon, S.Y., Hoffman, C.W.D., Gonzalez, A.J., Knauf, R., and Barr, V.B. 'Validation of Human Behavior Representations'. In *Foundations for V&V in the 21st Century Workshop*, 22–4, Laurel, MD, 2002.
- Hester, P.T., and Tolk, A. 2010. 'Applying Methods of the M&S Spectrum for Complex Systems Engineering'. Paper presented at the Emerging Applications of M&S in Industry and Academia Spring Simulation Multi-Conference, 17–24, 11–15 April, Orlando, FL.
- Hoffman, R.R., Crandall, B., and Shadbolt, N. 'Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis'. *Human Factors* 40/2 (1998): 254–76.

- Holland, J. H. 'Escaping Brittleness: The Possibilities of General Purpose Learning Algorithms Applied to Parallel Rule-based Systems'. In *Machine Learning: An Artificial Intelligence Approach*, Vol. II, edited by Michalski, Carbonell, and Mitchell, 593-623. Los Altos, CA: Morgan Kaufmann, 1986.
- Huang, H. M., Messina, E., and Albus, J. *Autonomy Level Specification for Intelligent Autonomous vehicles*. Gaithersburg, MD: National Institute of Standards and Technology, 2003.
- Jennings, N. R., Sycara, K., and Wooldridge, M. 'A Roadmap of Agent Research and Development'. *Autonomous Agents and Multi-Agent Systems* 1/1 (1998): 7-38.
- Kendoul, F. 2013. 'Towards a Unified framework for UAS Autonomy and Technology Readiness Assessment (ATRA)'. In *Autonomous Control Systems and Vehicles: Intelligent Unmanned Systems*, edited by Nonami, K., Kartidjo, M., Yoon, K.-J., and Budiyo, A. 55-72. Tokyo: Springer, 2013.
- Maier, M. W. 'The Role of Modeling and Simulation in System-of-Systems Development'. In *Modeling and Simulation Support of System-of-Systems Engineering Applications*, edited by Rainey, L. and Tolk, A., 11-41. Hoboken, NJ: Wiley, 2014.
- Matarić, M. J. *The Robotics Primer*. Cambridge, MA: MIT Press, 2007.
- McKee, L. 1998. 'Latency and Its Effects on the Fidelity of Air-to-Air Missile T&E Using Advanced Distributed Simulations'. *Joint Advanced Distribution Simulation/Joint Test and Evaluation*, Albuquerque, NM.
- Moya, L. J., and Tolk, A. 'Towards a Taxonomy of Agents and Multi-agent Systems'. In *Proceedings of the 2007 Spring Simulation Multi-Conference-Volume 2*, 11-18. Society for Computer Simulation International, San Diego CA, 2007.
- National Research Council (NRC). *Autonomy Research for Civil Aviation: Toward a New Era of Flight*. Aeronautics and Space Engineering Board. Washington, DC: National Academic Press, 2014.
- Neugebauer, E., Nitsch, D., and Henne, O. *Architecture for a Distributed Integrated Test Bed*. NATO RTO Modelling and Simulation Group Symposium (MSG-069), RTO-MP-MSG-069, paper 19. Brussels, 2009.

- North Atlantic Treaty Organization (NATO). *NATO Modelling and Simulation Master Plan – NATO Modelling and Simulation Implementation Plan Version 2.0*. Document AC/323/NMSG(2012)-016, Brussels, 2012.
- Ören, T. I., Ghassem-Aghaee, N., and Yilmaz, L. ‘An ontology-based dictionary of understanding as a basis for software agents with understanding abilities.’ In *Proceedings of the 2007 Spring Simulation Multi-Conference-Volume 2*, 19–27. Society for Computer Simulation International, San Diego CA, 2007.
- Page, E. H., and Lunceford, W. H. ‘Architectural Principles for the US Army's Simulation and Modeling for Acquisition, Requirements and Training (SMART) Initiative.’ *Proceedings of the Winter Simulation Conference*, IEEE, Piscataway, NJ, 2 (2001): 767–70.
- Parsons, S., and Wooldridge, M. ‘Game Theory and Decision Theory in Multi-agent Systems’. *Autonomous Agents and Multi-Agent Systems* 5/3 (2002): 243–54.
- Schelling, T. C. ‘Models of Segregation’. *The American Economic Review* 59/2 (1969): 488–93.
- Siegwart, R., Nourbakhsh, I. R., and Scaramuzza, D. *Introduction to Autonomous Mobile Robots*. 2nd edition. Cambridge, MA: MIT Press, 2011.
- Tisue, S., and Wilensky, U. ‘Netlogo: A Simple Environment for Modeling Complexity’. In *International Conference on Complex Systems*, 16–21, 2004.
- Tolk, A., and Uhrmacher, A.M. ‘Agents: Agenthood, Agent Architectures, and Agent Taxonomies’. In *Agent-Directed Simulation and Systems Engineering*, edited by Yilmaz and Ören, 75–109. Weinheim, Germany: Wiley-VCH, 2009.
- Tolk, A. ‘Modeling Sensing’. In *Engineering Principles of Combat Modeling and Distributed Simulation*, edited by Tolk, A., 127–44. Hoboken, NJ: Wiley, 2012.
- Tolk, A., and Rainey, L.B. ‘Toward a Research Agenda for M&S Support of System-of-Systems Engineering’. In *Modeling and Simulation Support of System-of-Systems Engineering Applications*, edited by Rainey and Tolk, 583–92. Hoboken, NJ: Wiley, 2014 .
- Wooldridge, M. J. *Reasoning about Rational Agents*. Cambridge, MA: MIT Press, 2000.

- Yilmaz, L. and T. Ören. *Agent-Directed Simulation and Systems Engineering*.
Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2009.
- Zeigler, B.P. 'Toward a Simulation Methodology for Variable Structure Modeling'.
In *Modeling and Simulation Methodology in the Artificial Intelligence Era*,
edited by Elzas, M.S., Oren, T.I., and Zeigler, B.P., 195–210. Amsterdam:
North Holland, 1986.

Afterword

Looking Forward – A Research Agenda

One of the key strengths of NATO is its ability to bring together diverse contributions of national experts for the collective benefit of the 28 members of the Alliance. The role of the NATO Science & Technology Organization (STO) is to leverage and augment science and technology capacities and programmes to contribute to NATO's ability to influence security and defence related capability development. Therefore, as the NATO Chief Scientist and Chairman of the NATO Science & Technology Board (STB), I welcome initiatives such as this volume that contribute to collaborative science, research, and technology development on critical capability issues facing the Alliance. And in this case, I'm pleased to welcome contributions by experts from NATO partner nations of Austria, Finland, and Switzerland.

The works in this volume are complimentary to the ongoing work in the various Panels and the Centre for Maritime Research & Experimentation (CMRE) under the STB, which has defined *autonomy* as one of the S&T Areas in the NATO Science & Technology (S&T) Priorities. This volume aligns with the STB view of a multidisciplinary approach, as was demonstrated during the 2014 STB Symposium on Autonomous Systems. The STB Symposium clearly identified that the concept of autonomous systems is not an isolated discipline in itself, but as an end result of a constellation of supporting and precursor technologies and capabilities. It is pertinent to note that other S&T Areas in the NATO S&T Priorities also need signification progress in order to achieve autonomy: *advanced human performance, cultural, social & organisational behaviours, data collection & processing, information analysis & decision support, communications & networks, power & energy, and advanced system concepts*. Research and development programs concerning autonomous systems need to consider these supporting S&T Areas both separately and in integrated programs to maximise capability development potential.

With this in mind, the following broad areas would benefit from particular consideration by the S&T, policy and operational communities in an integrated fashion.

Operational use of autonomous systems

Autonomous systems present a tremendous opportunity for the development of novel operational concepts such as swarming and human-machine teaming. These new concepts may open up a wide array of hitherto unanticipated ways of operating. This is both an opportunity and a threat: while NATO and Alliance members will certainly seek the advantages of such operational concepts, adversaries will also seek their gain, and we must be ready. The S&T Area of *advanced system concepts* addresses these ideas. There are many opportunities for studies and analysis of how warfighting will be affected: simulations of swarms, war-gaming, and optimization and effectiveness of manned-unmanned team combinations.

With the integration of autonomous elements, the way in which systems are tested for operational use will have to change. Trust, both in autonomous systems and their use, was also identified during the aforementioned STB symposium as a key area for future research. To support this, the concepts of system verification, validation, testing, and evaluation methods; reliability and certification standards; machine-to-machine interoperability; training systems; and auditing and control systems all need to be considered in the context of the safe and predictable operational use of autonomous systems.

Human – machine teams

At the core of the adoption and integration of systems with increasing degrees of autonomous behaviour is the concept of human-machine teams. As the technology capabilities improve, the relationship between the human and the machine will have to evolve, so that in the end both can work seamlessly together to deliver on commander's intent. At the recent STB Symposium on Autonomous Systems, the theme of manned-unmanned teams was identified as an area still requiring considerable S&T effort. In this respect, it is critical to involve the military operator in these research endeavours as military doctrine should co-evolve together with the changing technology landscape.

Impact on the character of military operations

While much of S&T covers the details of systems and technologies and the potential evolution of operational concepts, strategic-level research should be encouraged on the potential impact of autonomous systems on the nature and conduct of war. This is related to another S&T Area – *cultural, social and organisational behaviours*. The rising use of systems with an increasing degree of machine autonomy will likely change military structures and even the core of certain military professions. It is conceivable that in the future, aircraft pilots may eventually become “system managers;” the question remains as to the impact on military organizations, training, and culture. In the first chapter in this volume, Scharre rightfully notes that while the nature of the tasks and tools may change, the ethos embodied in military professions will remain.

Countering autonomous systems

It is likely that systems with autonomous capabilities will proliferate and will be found in the future operational environment. There are unique challenges that will be faced in countering these systems, especially without the ability to rely on traditional jamming and communication-link disruption methods. Policy makers are advised to encourage research, development, and planning to counter potential conventional and asymmetric uses of autonomous systems against friendly states.

A related issue to the counter-autonomous system challenge is that of cyber defence. A critical threat to autonomous systems could be their vulnerability to hijack by computer hackers. While the very fact that they operate with reduced communications partially prevents this, it also increases the risk, as recovery becomes more challenging. It is no coincidence that *communications & networks* is one of the S&T Areas in the NATO S&T Priorities.

Command and control

Finally, research is needed on the impact of autonomy on traditional command and control concepts. Autonomy will allow highly decentralized command centres located both in and outside theatre and more flexible and agile network control systems relying on machine-to-machine communications, rather than direct operator control. This may blur considerably the boundaries between command levels. Future research must focus on battle space management

concepts, and operational command. Again, this relates to the *cultural, social & organisational behaviours* S&T Area: military organizations are going to change and the Alliance must be ready.

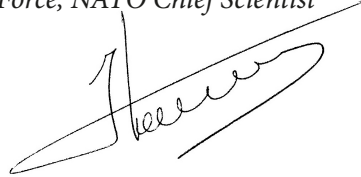
The way forward

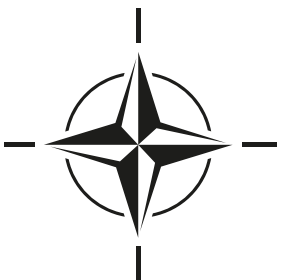
The military opportunities and challenges posed by the progression towards military systems with great autonomy are both great. S&T, as an enabler to military capability, has a significant role to play across the spectrum of capability development. This ranges from providing the strategic analysis necessary for military leadership to make the future capability decisions to the research and development necessary to enable the tactical and operational advances that will keep the our militaries at the forefront of military effectiveness.

And across all of these areas, the continuous and open dialogue and engagement between the S&T and military communities will be the underpinning of our success. In the words Theodore von Kármán:

“Scientific results cannot be used efficiently by soldiers who have no understanding for them, and scientists cannot produce results useful for warfare without an understanding of the operations.”

Major-General Husniaux, Belgian Air Force, NATO Chief Scientist

A handwritten signature in black ink, appearing to read 'Husniaux', with a long horizontal stroke extending to the right.



North Atlantic Treaty Organisation
Headquarters Supreme Allied Commander Transformation
7857 Blandy Road, Suite 100
Norfolk, Virginia. U.S.A.
23551